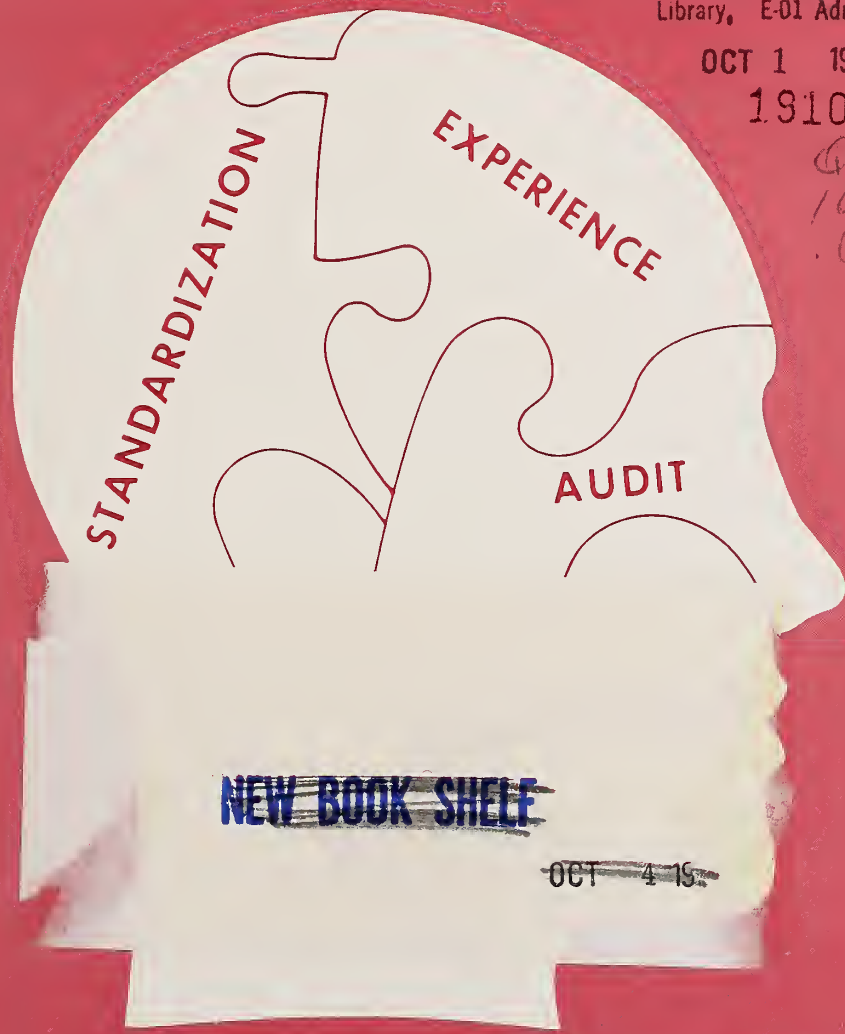
NATIONAL BUREAU OF STANDARDS

NBS SPECIAL PUBLICATION 451

U.S. DEPARTMENT OF COMMERCE / National Bureau of Standards



DATA BASE DIRECTIONS
The Next Steps

NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards[1] was established by an act of Congress March 3, 1901. The Bureau's overall goal is to strengthen and advance the Nation's science and technology and facilitate their effective application for public benefit. To this end, the Bureau conducts research and provides: (1) a basis for the Nation's physical measurement system, (2) scientific and technological services for industry and government, (3) a technical basis for equity in trade, and (4) technical services to promote public safety. The Bureau consists of the Institute for Basic Standards, the Institute for Materials Research, the Institute for Applied Technology, the Institute for Computer Sciences and Technology, and the Office for Information Programs.

THE INSTITUTE FOR BASIC STANDARDS provides the central basis within the United States of a complete and consistent system of physical measurement; coordinates that system with measurement systems of other nations; and furnishes essential services leading to accurate and uniform physical measurements throughout the Nation's scientific community, industry, and commerce. The Institute consists of the Office of Measurement Services, the Office of Radiation Measurement and the following Center and divisions:

Applied Mathematics — Electricity — Mechanics — Heat — Optical Physics — Center for Radiation Research: Nuclear Sciences; Applied Radiation — Laboratory Astrophysics[2] — Cryogenics[2] — Electromagnetics[2] — Time and Frequency[2].

THE INSTITUTE FOR MATERIALS RESEARCH conducts materials research leading to improved methods of measurement, standards, and data on the properties of well-characterized materials needed by industry, commerce, educational institutions, and Government; provides advisory and research services to other Government agencies; and develops, produces, and distributes standard reference materials. The Institute consists of the Office of Standard Reference Materials, the Office of Air and Water Measurement, and the following divisions:

Analytical Chemistry — Polymers — Metallurgy — Inorganic Materials — Reactor Radiation — Physical Chemistry.

THE INSTITUTE FOR APPLIED TECHNOLOGY provides technical services to promote the use of available technology and to facilitate technological innovation in industry and Government; cooperates with public and private organizations leading to the development of technological standards (including mandatory safety standards), codes and methods of test; and provides technical advice and services to Government agencies upon request. The Institute consists of the following divisions and Centers:

Standards Application and Analysis — Electronic Technology — Center for Consumer Product Technology: Product Systems Analysis; Product Engineering — Center for Building Technology: Structures, Materials, and Life Safety; Building Environment; Technical Evaluation and Application — Center for Fire Research: Fire Science; Fire Safety Engineering.

THE INSTITUTE FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides technical services designed to aid Government agencies in improving cost effectiveness in the conduct of their programs through the selection, acquisition, and effective utilization of automatic data processing equipment; and serves as the principal focus within the executive branch for the development of Federal standards for automatic data processing equipment, techniques, and computer languages. The Institute consists of the following divisions:

Computer Services — Systems and Software — Computer Systems Engineering — Information Technology.

THE OFFICE FOR INFORMATION PROGRAMS promotes optimum dissemination and accessibility of scientific information generated within NBS and other agencies of the Federal Government; promotes the development of the National Standard Reference Data System and a system of information analysis centers dealing with the broader aspects of the National Measurement System; provides appropriate services to ensure that the NBS staff has optimum accessibility to the scientific information of the world. The Office consists of the following organizational units:

Office of Standard Reference Data — Office of Information Activities — Office of Technical Publications — Library — Office of International Relations — Office of International Standards.

---

[1] Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D.C. 20234.

[2] Located at Boulder, Colorado 80302.

DATA BASE DIRECTIONS

The Next Steps + Special publication

Proceedings of the Workshop of the National Bureau of Standards and the Association for Computing Machinery, held at Fort Lauderdale, Florida, October 29-31, 1975

John L. Berg, Editor

Systems and Software Division
Institute for Computer Sciences and Technology
National Bureau of Standards
Washington, D.C. 20234

Richard G. Canning, General Chairperson

Working Panel Chairpersons:

Donald L. Adams, Robert W. Bemer,
George Dodd, R. Michael Gall, and
Charles D. Trigg

Sponsored by:

National Bureau of Standards

Association for Computing Machinery

U.S. DEPARTMENT OF COMMERCE, Elliot L. Richardson, *Secretary*

Edward O. Vetter, *Under Secretary*

Dr. Betsy Ancker-Johnson, *Assistant Secretary for Science and Technology*

NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Acting Director*

Issued September 1976

Library of Congress Cataloging in Publication Data
Main entry under title:
Data base directions, the next steps.
(National Bureau of Standards special publication ; 451)
Supt. of Docs. no.: C13.10:451
1. Data base Management--Congresses. I. Berg, John L. II. United States. National Bureau of Standards. III. Association for Computing Machinery. IV. Title. V. Series: United States. National Bureau of Standards. Special publication ; 451
QC100.U57 no. 451 [QA76.9.D3] 602'.1s
[001.6'442] 76-608219

National Bureau of Standards Special Publication 451

Nat. Bur. Stand. (U.S.), Spec. Publ. 451, 175 pages (Sept. 1976)

CODEN: XNBSAV

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1976

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402
(Order by SD Catalog No. C13.10:451). Stock No. 003-003-01662-4 Price: $2.40
(Add 25 percent additional for other than U.S. mailing).

FOREWORD

This report constitutes the results of a three-day workshop on data base systems held in Fort Lauderdale, Florida on October 29, 30 and 31, 1975. The workshop was sponsored jointly by the National Bureau of Standards (NBS) and the Association for Computing Machinery. The workshop continues the close working relationship that was started in 1972 between the Institute for Computer Sciences and Technology of NBS and a major professional organization, the ACM.

The idea to hold a workshop was proposed to NBS and ACM by Mr. Richard Canning who was appointed General Chairman. The purpose of the workshop was to bring together leading users, managers, designers, implementors and researchers in the area of data base technology to provide insight for managers facing data base management decisions.

John L. Berg, Editor

ABSTRACT

What information about data base technology does a manager need to make prudent decisions about using this new technology? To provide this information the National Bureau of Standards and the Association for Computing Machinery established a workshop of approximately 80 experts in five major subject areas. The five subject areas were auditing, evolving technology, government regulations, standards, and user experience. Each area prepared a report contained in these proceedings. The proceedings provide guidance of steps managers should follow to prepare themselves and their organization for the installation of data base management concepts. The auditing working panel noted the increased vulnerability of organizations who integrate their formerly dispersed and redundant files into a data base and suggest actions to address this risk. The technology report noted several promising parallel developments but concluded that the future would see evolving, rather than revolutionary data base progress. Government regulations, particularly the drive for individual privacy rights, were seen to play an important role in determining data base directions and the panel's guidance on cost impact suggest that organizations would experience reduced costs with data base technology. Standards pervaded all issues and were found necessary in several sub-areas of data base technology but the panel saw no immediate likelihood of national data base standards. The user experience working panel noted that data base systems had impacted their organizations to the extent of reconsidering existing data flows, areas of responsibilities, and procedures.

Key Words: Auditing; cost/benefit analysis; data base; data base management; DBMS; government regulation; management objectives; privacy; security; standards; technology assessment; user experience.

TABLE OF CONTENTS

CONTENTS - continued

A MANAGEMENT OVERVIEW

What information does a manager need to use data base technology? This question prompted a joint NBS and ACM workshop to which data base authorities were invited to answer this question. Keynote speaker, Daniel Magraw, spoke for managers when he identified eight areas of need. These were:

1. Establish management's DBMS objectives
2. Have realistic expectations
3. Organize for data base systems
4. Perform cost/benefit analyses
5. Plan the transition to data base systems
6. Provide data base training
7. Anticipate the privacy issue
8. Recognize DBMS security implications

The workshop divided into five working panels to consider DBMS development from the standpoint of auditing, government regulations, standards, technology, and user experience.

In the following summary, the main points made by these five working panels have been combined under the above eight areas of need, thus providing some of the guidance that Magraw sought. In addition, the issue of data base standards pervaded all eight needs and has been treated under a special heading.

While this summary indicates some range of opinion, the reader should refer to the full reports for a more comprehensive look at the different viewpoints.

1. Establish management's DBMS objectives

Managers should begin considering data base technology by preparing written statements of objectives and plans. The data base technology plan should include these sub-goals:

- o Determine management goals and define benefits sought in a plan that has top management approval.

- o Prepare a cost/benefit analysis (see 4 below). Use this analysis to seek top management commitment for data base plan and to enlist middle management support. Involve middle management in planning, implementation, and usage by identifying their data base advantages.

- o Develop a five year (or more) program:

(1) Utilize middle management participation to select high pay-off applications and data suitable for integration.

(2) Plan for data base compatibility with existing data use, (i.e., insure doing no harm to existing systems).

(3) Develop a standards plan and its application scope (see Standards below).

- o Establish a data base administration function and integrate with existing organization (see 3 below).

- o Select a DBMS system on the basis of suitability for immediate needs, benefits, and high probability of pay-off:

(1) Seek system stability by selecting a system offering flexibility, appropriate data independence, and established standards.

(2) Evaluate supporting tools provided with DBMS such as auditing software, performance measurement, manual and automated tuning, security provisions (see 8 below).

(3) Assess carefully risks associated with committing to the selected DBMS in terms of the costs to back out or select an alternative DBMS. Determine ability to convert data, programs, and skills to other DBMS.

- o Prepare a plan for transition to DBMS and gradual phasing to the degree of integration and central control desired (see 5 below).

- o Prepare a plan for training. Include all levels of management in plan. Involve users in planning applications, sharing of data elements, and assigning data element responsibilities.

- o Include in planning, provisions for appropriate controls to meet legislative or regulatory requirements (see 7 below).

- o Address auditing needs early by including internal auditors in design phases and making available auditing tools appropriate for external auditors.

Recognize that DBMS increases the need for close, effective auditing and also it makes auditing more difficult to do. Understand that the auditors will need to access data independent of the routine method of access.

- Prepare a continuous program for monitoring DBMS effectiveness against current management goals.

2. Have realistic expectations

- Assess two areas of potential benefits with objective and quantifiable dimensions: those derived from the installed DBMS (see 4 below) and those anticipated from advancements in data base technology.

- Do not delay immediate DBMS benefits while awaiting perfected systems. Technological progress over the next five to ten years most likely will advance in evolutionary rather than revolutionary steps. No big surprises lie around the corner.

- Look for and encourage technical developments in the following areas:

  (1) Tools to measure and improve DBMS performance. Current technology can provide such tools now, but user pressure will speed their incorporation into DBMS.

  (2) DBMS performance simulators. These offer a useful, inexpensive alternative to actual performance measurements but may become subjective in important examination areas.

- Though DBMS usually relieve hardware constraints, new hardware can affect the degree of physical independence from DBMS to DBMS. Few objective or quantifiable measures of this degree of variance exist. Evaluation is often subjective and intuitive. However, managers should watch for:

  (1) Major breakthroughs in storage device speed which would lessen the need for optimization and "tuning" required for achieving DBMS pay-offs in existing systems.

(2) Major breakthroughs in on-line storage capacity which would permit more flexible usage of performance measures (like audit trails) since they, too, could be stored on-line for computer use.

- Anticipate the impact of these possible software advances:

  (1) Future system enhancements which will gradually offer more automatic performance "tuning" with resultant economies. Current DBMS performance measuring and improving techniques are generally applied "manually."

  (2) Future DBMS which offer representation-independence, flexible data structures, search path optimization, high level query languages, or data translation. Such systems would protect existing investments in data collection to the extent data conversion is cost-effective.

- New system architectures are foreseeable in the next five years. These include:

  (1) Distributed data bases in which several distinct data collections appear to the user as one integrated data base. Such systems offer options on trade-offs on such issues as multiple copies of files, security, and rapid access.

  (2) Minicomputers offering data base function support to improve main computer processing in the same way minicomputers currently take on communication functions.

3. Organize for data base systems

- Develop data base administration function as required by the extent of centralized control and data integration intended. This function should include:

  (1) Responsibility for data base design, structure, standards, and integration.

  (2) Selection and control of data dictionary entries.

  (3) Security and integrity considerations.

  (4) Privacy requirements.

- o Include auditing needs in early design considerations since use of DBMS intensifies auditor concerns about system controls, operating procedures, and standards.

- o Assign responsibility for each data element in the data base. The more data elements are shared, the more needed are clear responsibility and accountability.

- o Re-evaluate enterprise organization in view of shared data needs and specific responsibilities for data collection and maintenance.

4. Perform cost/benefit analysis methodology

- o Recognize that, while direct costs (particularly machine room costs) will be readily determined, DBMS benefits may reflect more indirect quantities like information value and ease of access to data. However, DBMS also has non-hardware, non-software costs in such things as organization, structural, and disciplinary changes. Positive indicators for DBMS payoff are a need for: access to large volumes of data by a wide variety of users, complex or unpredictable queries, concurrent access of shared data, complex information processing, and high levels of integrity and security.

- o The first benefit managers will realize comes from the formal study of the organization's data needs, its flow, and the responsibilities associated with data collection and maintenance.

- o Few empirical measures of DBMS cost/benefits analysis exist to assist the manager. Identified benefits often seem subjective. These include:

  (1) Reduced costs in programming, programming modification, and data conversion,

  (2) Reduced data redundancy with resultant hardware and processing savings as well as improved data accuracy since the need to update several copies in parallel is eliminated,

  (3) Availability of computer power and data to users without special computer skills. However, some degree of training is always desirable.

- Inevitably data base systems must co-exist with other systems. The cost of such parallel structures must be considered as well as the cost of insuring compatibility among the various existing systems.

- Investigate simpler file management systems for their ability to provide the benefits sought before accepting DBMS's increased complexity and higher overhead costs. Few requirements can justify a one-copy DBMS. Review such decisions very carefully.

- Avoid the home-built DBMS. DBMS are expensive, difficult systems. They require special development skills. Construction times are measured in years and require a long term commitment to reach a pay-off. Custom built systems lose the benefit of common investment. System testing, certification for security and privacy, and tools for conversion to the next system would all be expenses borne alone.

- Examine carefully the expense of data independence features and weigh it against the benefits needed. Determine the proper degree of data independence from the enterprise needs and the system's anticipated stability.

- Managers should use auditors to determine and monitor the cost/benefits of DBMS systems.

5. <u>Plan the transition to data base system</u>

- Prepare for transition:

  (1) Precede any data integration with data standardization.

  (2) Insure the availability of adequate hardware/software.

  (3) Plan a step-by-step conversion of existing data collections and applications. Phasing should minimize risks at each step.

- Provide for the natural resistance of people to changed methods, loss of data ownership, and loss of data control. Show benefits to each individual and obtain middle management's commitment to counter staff's reluctance to change.

- Develop in parallel to DBMS a repertoire of measurement, simulator, benchmarking, auditing, and tuning tools. Collect empirical data on data base usage, tree lengths/depths, query rates, and update rates.

6. Provide data base training

- o Conduct training appropriate for each phase of development and implementation. Standard procedures, terminology, and practices will ease training. Include training for managers at all levels. Use this opportunity to sell the new technology and its advantages. Use training to prepare managers for such sociological factors as reluctance to change, sharing of data files, etc. Provide technical training in data base design, system implications, and possible future directions of data base technology.

7. Anticipate the privacy issue

Although the Federal Privacy Act of 1974 applied primarily to Federal agencies, managers should anticipate a major growth in privacy legislation--including an extension to private industry.

Privacy legislation applies to personal data collections, whether automated or not. DBMS offers opportunities for cost-savings by utilizing centralization of control to simplify compliance.

Managers will seek certification of DBMS compliance to meet privacy requirements. Auditors face this task reluctantly because of its difficulty.

The major aspects of existing privacy legislation are:

CONTROLS ON OPERATING PROCEDURES

An organization must:

- o Take precautions against natural hazards and other threats to the system and its data
- o Publish descriptions of its system
- o Establish procedures for responding to inquiries from individuals about their records and for settling complaints about their accuracy
- o Keep a log of all users of each person's records and the intent of that use
- o Make an individual responsible for the enforcement of privacy legislation
- o Ensure that data is both timely and accurate

USAGE CONTROLS

An organization must:

- o Inform a subject of the intended use of the data, and inform the subject if a new use becomes apparent (implications of this in a shared data environment)

- o Use data only for its stated purpose
- o Transfer data to a new system only with the permission of the subject, and only after ensuring that the privacy of the data will be adequately maintained in the new system

8. Recognize DBMS security implications

Recognize the increased vulnerability resulting from centralization and integration of corporate data assets. Understand the need for catastrophe planning, total testing, fail-safe mechanisms, and audit capabilities.

Select a DBMS with provisions for:

(1) restoring service after failure,
(2) restoring data content to some previously known good state,
(3) validating input and update functions, (central control of data definitions makes validating standards easier,)
(4) self-diagnosing; DBMS checks its own links or chains, etc.,
(5) producing control totals for validations outside the DBMS,
(6) logging security or integrity violations,
(7) producing audit trails. Note that operating system or DBMS logs may not be sufficient for auditors,
(8) establishing terminal security including restricted access to terminals and other remote entry points.

Standards

- o DBMS, by its very nature, forces managers to develop standards. The earliest decision facing the manager is the scope to which the standard should apply. Do the company needs to achieve DBMS benefits require standards at the DBMS site only, company wide standards, adherence to national standards, or even international standards?

- o Standards, if they are to provide confidence in data base content, meaning, and use, should address four areas: terminology, criteria for standards, components (such as language, data definition, etc.) and usage procedures.

- o Impact of standards is pervasive. They facilitate protection mechanisms and procedures, transitions from system to system, training, interchange of applications and data, and the introduction of new technology. However, standards can inhibit the use and development of new products. Weighing the advantages and disadvantages falls on the manager of each DBMS installation.

- o Standardization requires prepared detailed specifications, and established maintenance support. Currently, only the CODASYL data base capabilities in COBOL were submitted as a candidate for standards. However, no national or international standard in DBMS seems likely in the next five years.

Conclusion

Prudent managers will approach DBMS with clear, immediate, and concrete benefits in mind. The approach includes careful preparation of the organization, planned transition, and step-by-step protection of existing functions. Implementation of the DBMS will proceed with continuous monitoring and training to meet intended objectives. Monitoring will continue throughout the lifespan of the DBMS to develop real data to optimize this and future DBMS.

Richard Canning, Chairperson; Jean Sammet, President, ACM; Seymour Jeffery, NBS; and John L. Berg, Proceedings Editor.

1. INTRODUCTION

Richard G. Canning
General Chairman

Biographical Sketch

Richard G. Canning, editor and publisher of EDP ANALYZER, has been active in the computer field since the late 1940's, first as an electronics engineer at IBM, next in the Navy guided missile program, then on a UCLA research project, then in consulting, and most recently (since 1963) in publishing.

He was a member of the National Council of ACM for four years and is active in ACM professional development and special interest group activities. He was a member of the Board of Directors and an officer of AFIPS from 1968 to 1971. Currently, he is the AFIPS representative to the IFIP Applied Information Processing Group (IAG) and a member of that group's Board of Directors.

1.1 Motivation

The person who makes the fundamental decisions in an organization on using data base technology has a difficult task before him. In making those decisions, he wants to select the most effective course of action for the near term and at the same time not end up on a dead-end path a few years from now. Further, he must select that course of action from among many alternatives.

For instance, here are just a few of the questions that this decision maker faces:

* Should we be considering converting to a data base? Under what conditions is a data base almost necessary?

* Should we standardize on a particular data base management system (DBMS)? What is the outlook for a national standard DBMS?

* How can our auditors assure themselves that they are getting access to all appropriate records in a data base?

* Which now existing and proposed government regulations appear to have the most impact on costs and methods of use of data bases?

* Are any technical breakthroughs in DBM technology likely in the next five years that will clearly obsolete the current DBMS?

1.2 Goals

To help the data base decision maker on questions such as these, the National Bureau of Standards and the Association for Computing Machinery organized this working conference. Our plan was to invite some 60 to 70 people, from the U.S., Canada, and Western Europe, selected for their knowledge of the field. They would spend 2+ days, organized in five working panels, addressing the basic question: What is expected to happen in the data base area in the next five years? Each working panel was to consider this question from one of five aspects--auditing of data bases, the effect of government regulations on data bases, evolving data base technology, standards for DBMS, and projections based on user experiences with data base technology. Each working panel was to summarize its conclusions in a report. Together, these reports, plus some other material, would constitute the conference proceedings. These proceedings were to be published by the National Bureau of Standards so as to be available to data base decision makers.

1.3 Accomplishments

Our main work product, then, was to be this report that you are now reading.

The working conference was held according to plan, at the end of October, 1975. This is the report of our work. Incidentally, in it you will find some guidance on the questions posed above, plus a lot, lot more.

Of course, it will be for you, the reader, to judge how well we succeeded in providing you with practical guidance for the use of data base technology. We do not claim that this report will provide the answers to all of your questions. I do feel, however, that it will help remove some of the questions in your mind as to what is likely to happen in data base technology and usage in the next five years. The participants were knowledgeable in data base technology and its uses. This report captures the consensus of their thinking. I suspect that you, like myself, may find some surprises in what they have to say.

I will not attempt to give any highlights of the report; for that, please read the Management Overview.

If this report can provide you with some "fixes" on questions relating to effective use of data base technology in the next few years, we will have met our goal.

2. A MANAGER'S VIEWPOINT

Daniel B. Magraw
Keynote Speaker

Biographical Sketch

Daniel B. Magraw is the Assistant Commissioner for Management, Department of Administration, State of Minnesota. For the past eight years and until just recently, he has been responsible for all aspects of the State of Minnesota information systems activities. One of the founders of the National Association for State Information Systems, he is a past president and, currently, a member of the NASIS Finance and Executive Committees. His nearly thirty years' experience in systems activities is almost equally divided between the private and public sectors. He taught courses in Systems for 22 years for the University of Minnesota Extension Division. A frequent speaker on many matters relating to information systems, he has been deeply involved in the development of federal and Minnesota data security and privacy legislation. His present responsibilities include management coordination and improvement, program evaluation, and issue analysis for the Minnesota state government.

The workshop planners recognized the need for a keynote talk that would pull together the five working panels and many different personal viewpoints into one common program. That program would address the needs of a manager about to consider data base systems. Mr. Magraw presents the manager's viewpoint and needs so well that we present his talk in full.

Editor

2.1 INTRODUCTION

When Dick Canning called to ask me to speak to this highly select group, I was completely surprised, and said so. And then I asked: "Why me?" I can report to you that the only response I heard that made sense was that my initials were right --- DBM. I was raised in the midst of the alphabet soup days of FDR's New Deal and have lived through the latter day exponents of acronym based government; --- and now that my initials do stand for something of national consequence, I have a feeling of fulfillment.

Seriously, I am very happy to be here today at a conference which can and I expect will have long range influence. My view is that conferences on major issues, like this one, under the kind of sponsorship as we have here, do have an impact, particularly if well documented and especially if followed up with one or more subsequent meetings. And also seriously, I have had two major qualms about this task of keynoting: first, about the need for anyone to "keynote" a conference of workshops composed of many experts in various facets of this field, and second, about my own inadequacies. On the first point, perhaps there is some advantage to you to listen to someone who is at least a layman in the field as you shift gears from your regular activities preparatory to this three day stint. On the second point, I want to make it clear that I am not a data base expert (although I might qualify as one of DBM's leading cheer leaders). I finally concluded that hearing from a non-data base management expert but one who has been heavily involved both as a manager of computer professionals and as an executive involved both in providing and using information may be appropriate.

2.2 IS THERE A NEED FOR DBM GUIDELINES?

The objective of this conference, through its component workshops, is to develop a set of guidelines covering the principal aspects of DBMS. Perhaps we should ask whether guidelines are really needed in this field, or are NBS and ACM simply emphasizing their and our own parochial interests and attributing an importance to a technique way beyond what it deserves? Who needs them? Who wants them?

My emphatic answers are that management does need DBM guidelines, that management has needed them for some time, that it is impossible to overemphasize the importance of DBM to the management process, that NBS and ACM are to be commended for joint initiation of the meeting, and that I wish they had done it at least two years ago.

2.3 CENTRAL MANAGEMENT ROLE OF DATA

As support for those answers, let me suggest this. The essence of management, according to all authorities and to common sense, is rational decision making <u>based on the best available data</u>. Thus, the broad questions of data management lie at the very center of the management process. And they have always been there. But now we have techniques and equipment which make it possible to do much more than previously with data in terms of capturing, storing, editing, organizing, retrieving, relating, manipulating, analyzing. In short, we have the capability now of moving more rapidly toward obtaining maximum utility out of our data resources. When I say "we" in this context, I mean overall management, not EDP or DB management.

2.4 FUNDAMENTAL DIFFICULTY OF THE TASK

Now hear this! When anyone commences the serious business of rationalizing the processes related to data and begins working toward the design and implementation of an orderly, flexible, and comprehensive DBMS, he finds himself digging around in the very heart and soul of management. It is an exceedingly complex endeavor; and because of the nature of the subject matter, that is, data and information providing the raw material for managers' decision-making and for evaluation of their performance, it is also an exceedingly dangerous pastime.

2.5 MANAGEMENT NEED AND DESIRE FOR GUIDELINES

Perhaps we here can agree, then, that the top executive who is serious about DBMS can be materially aided by a set of DBMS guidelines drafted by a group of people experienced in all aspects of DBMS. We can also agree that the top executive who understands the promise and also the complexities and difficulties of DBMS will be eager for such guidelines. As a corollary to this, perhaps we also agree that if there is not some top executive direction of the process, nothing of DBMS consequence will happen --- guidelines or not. But more of that later.

2.6 LIMITATIONS ON DBM COVERAGE

Permit me to delimit the subject and to suggest that we do likewise in our workshops.

For many years, some of us have subscribed to the theory that all systems should and will be on-line, some sooner than others, but all eventually. Also, that computer files will be more cost effective than manual files --- for all purposes. My perception is that both of these theories are essentially sound and, if anything, are becoming realities faster than we have anticipated.

Many have also shared the dream that at some point, in whatever milieu we find ourselves, small organizations or large, private or public, all data will be available to us for browsing; and better yet, for converting to information of whatever type we ask; and far better than even that, for analyses resulting in formatted presentation to us of alternative decision packages with trade offs, accompanied by all appropriate probabilistic calculations.

I am not willing to give up that dream, nor do I think we have to give it up. On a limited basis, we are there in some sub-systems. And it is coming generally and with increasing speed.

But I would like the luxury of delimiting what we are talking about --- without having to tell you now or ever how I would specifically define those limits. Generally, I suggest we think in terms of data that are susceptible to one of two uses: an operational purpose that is typically but not necessarily repetitive; and a management purpose such as in forecasting, planning, command, control, and evaluation. The first group might be characterized by production type data for day-to-day purposes and the second by the term browse-worthy. What do these groups exclude? I said I would not answer that question. But I will say that in the context of having computerized all data in whatever form and of whatever nature (correspondence files, blueprints, etc.), then I think it excludes a bunch.

2.7 THE MANAGEMENT PROBLEMS

Now moving along quickly to get away from that delimiting exercise, I would like to express some of my views of the broad problems we are facing and thus indirectly of where I perceive the need for guidelines. These are covered in my general order of priority and are inter-related to varying degrees.

You will note that I am more concerned about shortcomings of the providers and the users than I am about the vendor's defects. By and large the history of computerization is that users lag behind the vendors, often far behind. The situation is best exemplified by that hoary with age story about the Agricultural Extension Agent. He was extolling to a Midwest farmer the virtues and advantages of taking courses in the Ag Extension's continuing education program. After brief consideration, the farmer demurred, saying, "I ain't farming nearly as well as I already know how."

I am going now to suggest eight areas of concern to me, the first one at some length, and the other seven more briefly. Then I am going to list 15 others.

2.7.1 PROBLEM 1. DBMS OBJECTIVES. The first problem is that there needs to be a clear and highly specific understanding of the objectives of DBMS in any organization. It may have been fashionable to keep up with the Joneses and install a computer or two. But 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. My own belief is that there is a sine qua non of such objectives --- and using a keynoter's prerogative, may I discuss it for a moment.

It was about 20 years ago when enough IBM 650's were installed so that even doubting Thomases could see the handwriting on the wall. I have to subtract 1955 from 1975 on paper and then with a calculator. I can't believe it is only 20 years --- at least when I look at the state of the hardware, software, and communications art, and when I try to conceive of the overall impact of computers on society.

If there is a use for the word "incredible", it would be for me, at least, to describe the phenomenon of the exponential rate of increase of that impact --- and of its continuation.

But what about the history of decision-making during that time? Almost from the beginning of that period and still today, there has been a constant refrain in the background going something like this: "Computers are good for cost reduction in many routine areas. But you ain't seen nothing yet." Followed by a "Wait until next year!" shout. We know about that in Minnesota with our Vikings.

Without any real quantification behind it, my view is that we really "ain't seen nothing yet" in terms of computer decision-making (and I don't mean simply in the decision-making process although that obviously is important). I say this even though I have spoken ad nauseam on behalf of the computer as a decision maker. Literally billions of decisions, formerly made by humans in trivial areas, are made daily by computers. And thousands of fairly profound decisions (like cracking plant production decisions) are made daily by computers.

And yet, many of us, both theoreticians and practitioners, are forced to say, "Wait until next year. Then decision-makers will really use computers for decision-making on a wide scale in most areas of management."

Why are we still saying that? The answer lies not in shortcomings of the computer manufacturers, of the operating software, of specialized data base software, or of systems designers. The answer lies in failure of management to define their decision systems so that data and information systems requirements for those decision systems can be addressed. This is a repeated and long-standing failure of management. Sure, it's tough. But decisions are indeed made. They are generally based on data and on methodical, logical analyses of the data (based on intuition normally only in the absence of data). And when data needs are known and can be satisfied and when the methodical, logical system for data analysis is known, decision-making can be aided by computers or can be partially or fully computerized.

But too often, management stands pointing at the industry or its own systems staff and attributing its own failure to them. It puts one in mind rather painfully of our exalted Congressmen shouting and pointing at the private sector, both labor and management, accusing them of causing inflation at a time when Congress is creating a deficit of between $70 - $100 billion a year and is going to a $2.00 bill in recognition of the fact.

The point is that DBMS objectives must be unequivocally established by top management. So often they are not. We all know of organizations in which the EDP'ers are madly pursuing DBMS with literally or virtually no management understanding or interest. And when the objectives are established, the DBMS programs and policies to reach those objectives must be strongly and visibly supported by management.

2.7.2 PROBLEM 2. REALISM. The second major problem which I see is related to the first: the question of realism. I expect it will be touched one way or another by each workshop. There is an enormous place entitled "Great Expections Cemetary" that is figuratively full of the bodies and literally full of the broken, altered, or shortened careers of professionals from all parts of the computer consortium. I have visited with myself there two times. Some arrived as a result of a small minority of vendors, academics, writers, providers, or users overselling, overrating, overstating, etc.

In DBM, we again are faced with that problem. And this time probably more acutely than ever. I say "probably" because many DBM failures are being noticed and they may have a sobering effect. We are also faced with another "expectation" problem related to DBM --- mini-computers. The song is being sung in many quarters, and with increasing gusto, that all you really need to do is set up each system or subsystem on a mini. Then you will save enormous sums and improve service to management. And it can be done virtually overnight. No problem on data integration. Simply hook onto the communication system and browse contentedly through all files, no matter where located. No problem of any sort. Amen!

But these are problems of realism for which we need -- I say desperately need -- guidelines.

2.7.3 PROBLEM 3. ORGANIZING FOR DBM. A third problem in management's baliwick relates to organizing for Data Base Management. DBM turns out to be an expensive, academic exercise unless an unequivocacal statement of responsibility and a proper structure are established for the function. It is not my task, and it would be presumptuous of me, to specify what the precise structure should be. But I cannot refrain from saying that unless there is strong, central, total authority over data item authorization and definition and all related DBMS functions, forget it!! That is, forget the data base concept. Further, the reporting relationships need careful attention. My not so surprising observation is that where DBM has been established as a separate function, reporting to the person in charge of information systems, DBMS progress is greater than where DBM has been assigned to a systems or technical support group. And with necessary redundancy, let me say that these guidelines must speak to the need for strong and visible management support for DBMS.

2.7.4 PROBLEM 4. DBM COST/BENEFIT. A fourth major problem that needs to be surfaced in a clearer manner for management is the cost benefit of DBM. On the cost side of the formula, I have particular concern for the non-hardware, non-software costs. What are the net costs of organizational, structural, and disciplinary changes necessary? What are the full costs of rationalizing the existing data base structures so that they are susceptible to DBM --- the cost of compatible data bases "painfully constructed piece by piece" as stated by James Martin in the Preface to his book entitled "Computer Data Base Organization." In our case (State of Minnesota), I judge that these costs will be enormous and the amortization thereof a slowly accelerating process. In any event, management needs checklists or guidelines to assure that it considers all significant factors of either cost or benefit.

2.7.5 PROBLEM 5. TRANSITION. A fifth problem is related to most of the previous ones, and I plead for advance forgiveness for redundancy in my own DBM. But I am simply flabbergasted at the cavalier dismissal, by friend and foe alike, of the overwhelming prospect of putting data bases in order for a DBM. Getting from where we are to where we want to be is not really a hardware-software problem. It is basically a data reorganization problem. (I am assuming prior and appropriate resolution of the organization structure and DBM authority questions.) From where I sit, there is entirely too much talk about the DBMS and related hardware and far too little about the basic problems of data definitions and organization, and the conversions and transitions associated with change. Where those are not being addressed, DBM is, and will continue to be, an acronym, not for Data Base Management, but rather for Data Base Mess.

2.7.6 PROBLEM 6. DBM TRAINING. A sixth area of concern to me is DBM training, particularly among the users. My understanding is that excellent training exists for computer professionals in DBM theory. I judge that there is insufficient training for them in the fields of decision theory and systems and in man-computer dialogue. But by far the largest and most urgent problem is in user training. This time, hopefully and almost prayerfully, I trust that the lip service given by top management of user and provider organizations to the need for their own training --- that the lip service will be replaced this time by a heavy dose of real training. This training can have a salutary effect on the realism problem discussed above --- and for that reason if for no other deserves prompt action. My presumption is, of course, that output from this workshop, and perhaps successor workshops, will be integrated into this training.

2.7.7 PROBLEM 7. DBM AND PRIVACY. A seventh item of concern to me is the privacy consideration, perhaps because of my heavy involvement in data privacy matters. We should be particularly watchful of such legislation insofar as it (1) restricts ability to interrelate data on individuals from various sub-systems and (2) limits data collection. Much privacy legislation is under consideration by Congress and state legislatures. Certainly, privacy implications call for some sharp restrictions on use of interrelated data but not outright prohibition on data interrelation.

Limitations on data collection will inevitably occur, I believe, on a broad scale because of language as in the Minnesota law which limits collection of data to that necessary for "administration and management of programs specifically authorized by the legislature," etc.

An interesting part of the Minnesota law requires that summary data be made available to anyone. No longer can any non-Federal public official in Minnesota say: "This summary data is mine, all mine. Stay away." This certainly bodes well for statewide DBMS, at least on summary data.

2.7.8 PROBLEM 8. DBMS SECURITY IMPLICATIONS. An eighth problem of great interest to us revolves around the security implications of DBMS's. Can we justifiably expect greater security capability? Do we become more vulnerable to some catastrophic event? Do we become indefensibly dependent upon one or a few people who are "all powerful" in a data sense and beyond effective control? Do we have reason to expect all inclusive, totally tested, fail-safe audit capabilities? Can we meet privacy criteria better through such systems?

Those are the eight biggest problems in my view --- and obviously skewed heavily toward user management understanding and involvement.

2.7.9 "MISCELLANEOUS" PROBLEMS. Some other questions that have been of concern to us are under what circumstances, if any, would an organization write its own data base software; what are the likely changes in the state of the art, both hardware and software, as related to DBMS, obviously of great importance; what considerations would be important in determining whether two different systems, such as TOTAL and SYSTEM 2000, would be appropriate in an organization; what implications for our DBM activities will result from developments in data dictionary/directory extensions; how close are we to having standard definitions in this field; are there any rules of thumb on costs of transition to DBMS of existing systems; are there any measures, based on actual experience, as to the effect of DBMS on data redundancy; is there any information indicating under what conditions redundancy is likely to be cost-effective; what information can be brought to bear on the question of building a DBMS containing a broad array of summary data; are there any indications of what are reasonable degrees of physical data independence or logical data independence; how does one assess risks inherent in adopting a proprietary DBMS marketed by other than a major factor in the industry; are there quantitative benchmarks of any sort that would be useful in controlling DBMS progress (data definitions, etc.); as an aid in setting priorities, what types of systems give evidence of quickest and/or greatest payoffs when initiated or converted to DBMS; as a corollary to the above, will it cost more or less to develop it under DBMS or not under DBMS; how seriously should we view the potential of the relational data base approach? What are implications of later shifting to that approach from the presently more conventional DBMS. This is by no means an exhaustive list of our concerns, let alone a composite of everyone's concerns. But let's hope it at least hints at all of them.

2.8 IN THE NEXT FIVE YEARS?

What really can we expect in the next five years? Such prognostications are, indeed, what is expected of us in these next three days, perhaps couched in part in terms of guidelines as to how to react to these prognostications. But permit me one prognostication as an input. Much is going to happen. Even if it is largely the rationalization of data bases to do more efficiently what we have been doing. Even if many or most never go to full-scale, all-out DBMS. Even if the hardware-software vis-a-vis DBM are behind schedule.

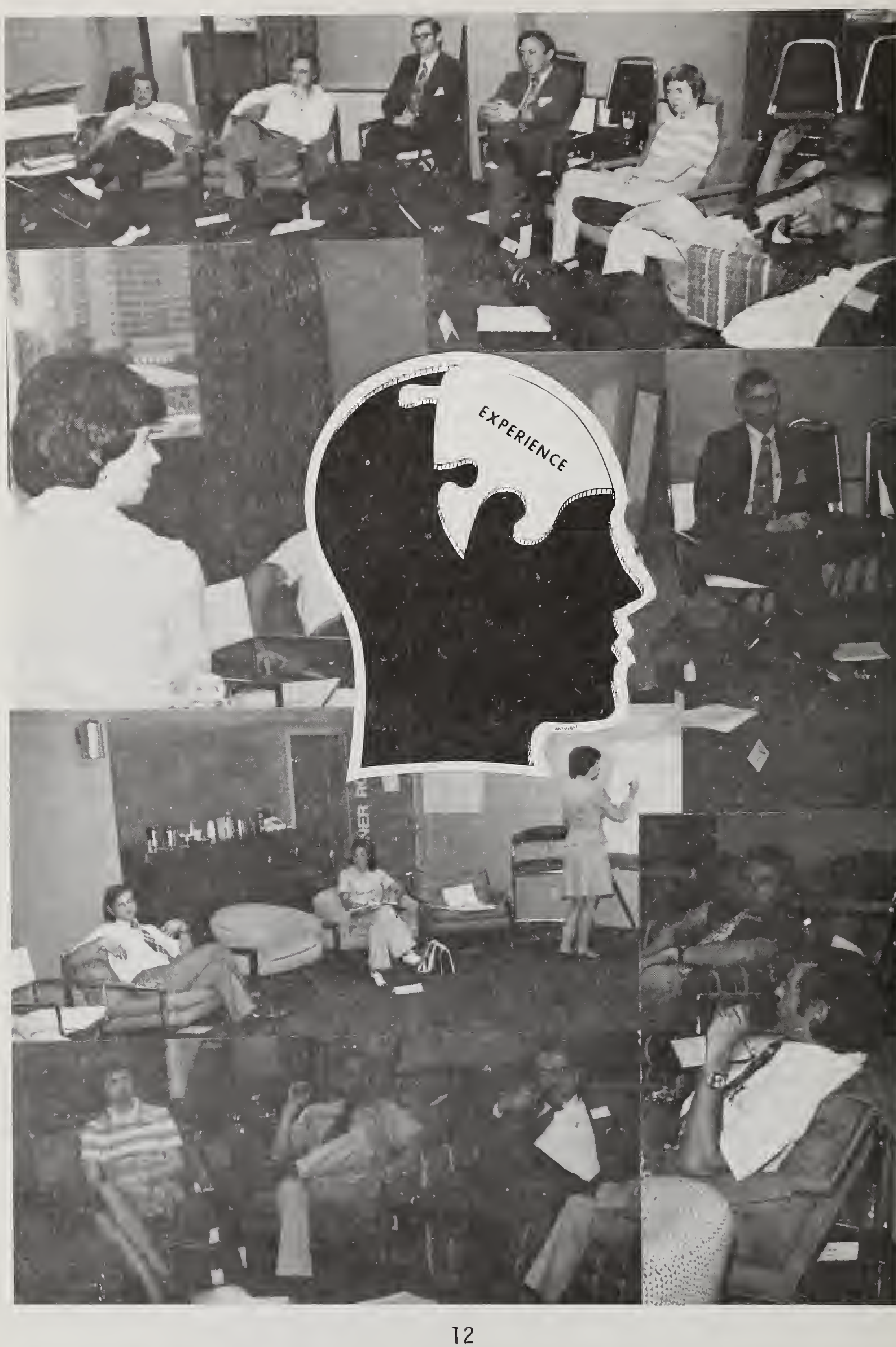
But the real pay-off will not be realized unless we move the DBM's into the area of decision system: aiding the decision-making process and actual decision-making.

And here there may not be much room for optimism. The requirements on management for their leadership and for their extended and detailed analysis and subsequent rationalization of their own decision systems are extremely onerous: onerous in terms of time and most of all in terms of the psychological impact of being not able to, or even if able to, spell out their own decision processes. Perhaps a review of what happened to the great promise associated with the phrase, and the practitioners, of Operations Research, would be constructive. That review might shed some light on what DBM guidelines should be.

In any event, those of us who have been saying, "Wait until next year" will be saying so at least for several years. And perhaps the best we can hope for is that there will be fewer of us in that group five years hence. But there will be basic progress in those five years, probably spectacular progress, from the combined efforts of hardware and software professionals and the data base administration and systems personnel.

As a specific output from this conference, however, I trust there will be a document that will provide the top executive who is seriously concerned about this problem a solid framework within which he can decide with some confidence what to do about DBMS.

And at least to that extent we will each have a part in bringing reality to James Martin's prognostication: "In centuries hence, historians will look back to the coming of computer data banks and their associated facilities as a step which changed the nature of the evolution of society, perhaps eventually having a greater effect on the human condition than even the invention of the printing press."


EXPERIENCE

3. WHAT EXPERIENCE HAS TAUGHT US

Working Panel Report On User Experience

Chairman: R. Michael Gall

Biographical Sketch

R. Michael Gall is the Associate Director for Information Systems of the Bureau of Manpower Information Systems at the U.S. Civil Service Commission in Washington, D.C. He has been in the Federal Government for 20 years, 10 of which have involved information processing-related assignments in Southeast Asia, Honolulu, California, and D.C. He holds a Bachelor of Civil Engineering from George Washington University, and a Master of Science in Information Science from the Georgia Institute of Technology. He is a graduate of the Federal Executive Institute and was a participant in the Government-wide Federal Executive Development Program.

Participants*

Margaret Derby, Recorder	Joseph Iwanski
Richard Duckett	Roger J. Kelly
Thomas Duff	Robert S. Korfhage
Ruth F. Dyke	Lowell S. Schneider
Elizabeth Fong	Victor A. Vella

3.1 Overview

The charter of the User Experience working group was to discuss and document experience related to the development, analysis, installation, and use of data base technology, and to extrapolate that experience for the next five years. The purpose of this effort was to provide a hypothetical Director of Information Services of an organization with pragmatic guidelines that could help him through the maze of conflicting information concerning data base technology.

The User Experience working group, comprising a diversity of experiences and points of view, achieved remarkable consensus almost consistently throughout the workshop. However, the most remarkable insights were (as one might expect) the most simple:

o Data bases have obviously been with us for a long time and have been managed in one fashion or another, some of them remarkably well, long before the advent of what we now call data base management systems and other related but as yet undisciplined terms.

---

* Complete addresses and affiliations are in Appendix C

o The most significant benefit of data base management dedication and commitment is the discipline that arises out of the effort and the organization to support it.

o Future technology is not likely to obsolesce any significant portion of dedicated effort toward better and more effective management of a data base invested at this time, and therefore, further delays while waiting for perfected tools should be avoided.

o Unless your needs are absolutely unique (which is unlikely), do not write data base management software in-house.

3.2 The Decision "to go data base."

It became eminently clear that once again we in the data processing profession have almost calculatingly given an aura of impenetrable complication to an essentially straight-forward function--the management of a data base. Out of this complication seems to be arising enough common descriptors and measures to begin to qualify data base technology as a discipline. What is confusing the issue is that this discipline embodies many segments of information processing previously vested in other disciplines of the profession. There was also felt to be an overreaction to implications of standardization, avoidance of cost-effectiveness analysis, inadequate education and training, and a great fear of a commitment to "go data base" on the part of most users. The group consensus clearly contained a central theme that "going data base" is only an evolutionary step along a path which has been followed for many years, and involves the application of new and diverse tools to make the data management function more effective, more productive, and more reliable than has been the case for the past decade.

3.3 Factors Affecting the Decision.

An initial questionnaire was distributed to the members of the working group and some of the responses to the questions (with some editorial commentary) are as follows:

Question 1. Is data base technology necessary or desirable for my organization? Is there a break-even point in size, type or complexity of information processing needs? If so, how do I measure it? Is there a way to determine cost/benefit?

Every organization already employs at least some concepts from data base technology.

Any organization that maintains information for subsequent access cannot help but use some concepts of data base technology. Even if their present techniques include only sequential access of tape records (or file jackets), these belong to the branch of computer/information science concerned with the structure, storage, and retrieval of information. But in recent years, this technology has evolved a body of new concepts and terminology aimed at synthesizing the entire discipline into a global perspective more conducive to formal study. This perspective, in which various strata of structure are visualized, and its attendant consequences, such as integration, is the new technology in the discipline about which everyone is concerned.

Several comments were offered which were considerably more specific:

If your organization needs three or more of the following, then data base technology is necessary for you:

a. An integrated data environment.
b. Rapid retrieval of data from large files.
c. A query/update language for use at terminals.
d. Backup and recovery requirements.
e. Privacy/security protection of your data.
f. Complex data structures.

And there were some specific "good news" reports:

In the case where the proper functioning of an operational unit is dependent upon complete and consistent data availability, the answer is yes. The function of the writer's office prior to the introduction of a DBMS was based upon single item scrutiny, memory and much manual searching. Via introduction of this formal DBMS approach, the level of productivity, accuracy and decision making has been enhanced.

Cautious approaches were very evident:

There is a break even point but it is difficult to define. Items that should be considered primarily are those relating to the user:

What is the desired goal?
What is the impact on the user's environment?
Will greater data availability improve or impede increased productivity?

And:

a. Size. The larger the file, the more likely you are to need data base technology.

b. Type. The type of query to be run is a factor. The more unpredictable the query, or the more complex the query, the more likely you are to need data base technology. The type of data structure is also a factor, as given below.

c. Complexity. The complexity of the data structure is an important consideration. Simple sequential or indexed sequential data files in fixed format have much less need for data base technology than those with complex structures or relationships or variable formats.

And, finally:

Complex information processing needs, not complex data structures, were the motivation for the construction of recent data base management systems. Complexity of processing can be assumed if the following requirements exist:

- fast access to large volumes of data
- by a wide variety of users (novice, semi-skilled, etc.)
- for the purpose of obtaining both anticipated and ad hoc information
- without compromise to data security or data integrity
- together with concurrent access to the same data by batch programs

Even a subset of the above conditions might be considered sufficiently complex processing needs.

It was clear, however, that "data base technology" had significant benefits not directly related to the improvement of processing. Certainly that aspect of data base technology which is concerned with data structure analysis and data structure notation applies to any organization. Data structure notation, such as Bachman Diagrams, are useful to the data processing professional in gaining an understanding of the nature of the data that is to be collected and processed. Differentiating between data elements that are hierarchically related and those which are network related...documenting these relationships in a way that promotes visual comprehension...and utilizing this "minds-eye-view" of the structured data to verify information processing requirements...these are the essential first steps in the design of any system.

Aside from the benefits to the designer in establishing a personal understanding of the nature of the data to be processed, data structure diagrams are useful in communicating with others. Data structure diagrams are easily understood (with a little tutoring) by non-professional data processors and thus make it feasible for the ultimate users of data to review and criticize proposed data base designs. User (especially management) participation in the development of information systems is always an important factor in the success or failure of a system; however, participation at a meaningful level requires an understanding of the problem and the consequences of alternative solutions. Structure diagrams provide a common language for communication between technician and user and, properly utilized, minimize the probability of a misunderstanding of a systems information potential.

A dominant theme which appeared several times was the benefit of a shift in viewpoint of data as a representation-independent model:

An organization can apply the new data base technology in more ways than just DBMS implementation.

The ability to view information on an abstract plane that is independent of its representation gives rise to several potential applications that could be beneficial to an organization. At the minimum, an organization could develop a representation-independent model of its information environment (via relations or whatever) purely to gain insight. This insight could be applied to decisions regarding management structure, information needs and information utility. Going one step farther, an organization could model alternative implementations of its information to assist in planning its information processing facilities. The candidates might well include some DBM's, but, more importantly, could also include improved file management techniques for their present application program environment, and even improved manual methods for their non-automated data. Finally, of course, an organization could actually implement the concept of structure separation either via an available DBMS, via in-house software or by a combination of both.

The measurable parameters, and particularly the problem of determining cost-benefit ratios, proved to be the usual problem. In some measure, it was more complicated than had been the case in data processing environments not specifically involved in overt data base technology:

Direct costs can be estimated with a great degree of accuracy. However, it is very difficult to append any accurate value to the benefits devised from improved productivity and/or performance. A subjective decision by management as to whether the anticipated improvement warrants the expenditure on both the short and long term is required. Careful planning and presentation can simplify the decision making process...you should be able to devise general parameters for measuring the impact of DBMS in an organization. Our organization has devised some key benefits from this approach. For example, a request for payment can be viewed, not by itself any longer, but in the context of the total funding sources, contract progress, specific legislation or other data relating to an item via a series of pre-formatted inquiries of various complexities.

Typically, non-specific factors predominate in the analysis:

It would be very difficult to construct a meaningful cost/ benefit analysis which would be a decisive factor in making this decision. Rather, the decision will probably be based on factors other than cost, such as data integration, on-line retrieval capability, complex data structure, etc.

But the consensus of the group evolved about the extended scope of the newly emerging issues of information value and the quantification of the worth of access:

Quantifying the cost-effectivity of data base technology applications is a formidable task.

If we limit the sphere of analysis to just the computer room, determining the cost-effectiveness of any of these applications is difficult but possible; difficult because the figures-of-merit regarding information systems are trade-offs (retrieval speed vs update speed; precision vs recall, etc); possible because all we need are some a priori objectives and some imaginative mathematical techniques. But the real problem is not limited to computer system performance. To get a true answer, we have to include the performance of the human system as well. And if we extend the sphere of analysis to include these factors, we invariably find ourselves faced with quantifying the value of information, and the value of information accessibility. These tasks are complete research topics in themselves. The only segment of the industry, to my knowledge, to have explored these topics is the defense/intelligence community and their answers, if any, are garrisoned in impenetrable vaults. If this workshop could accomplish no more than a guideline for future investigation of this area, it would be, in my opinion, an unqualified success.

Question 2. What is included in my database? How can I achieve integration of current files, text, graphics, random data, etc?

The pragmatic views dealt with specific problems faced by users; and stressed careful, planned, phased efforts:

The data base contains whatever data elements that are required to satisfy the information needs of the applications which reference the data base.

Past experience seems to indicate that phased data base implementations are easier to control and thus, more likely to succeed. Consequently, it is important that a data base management system be designed to minimize the impact on existing applications when the structure of a data base is extended to support new applications.

One basic advantage of integrating existing files, so that they can be processed by a data base management system, is to achieve centralized control of data. Centralized controls means greater security and enhanced data integrity. If greater security and/or integrity is required then there is a definite incentive to give priority to the integration of such data.

Another incentive for file integration is to take advantage of the more sophisticated data structuring capabilities of a data base management system. Applications systems which process network related data can generally be simplified by converting existing files to a data base file organization.

The order of activities was felt to be extremely important:

Several factors must be present for a successful conversion effort.

1. A plan for the total integration effort. This plan will cover approximately five years, and must be comprehensive enough to serve as an umbrella under which short-term plans are made.

2. Data standardization must precede data integration.

3. Adequate hardware and software must be procured.

4. A step-by-step conversion plan must be developed.

A data base is a model of the real world.

The business enviroment of any organization is the collection of entities that cause, result from, or in any way influence the organization's activities. As we cannot economically keep track of the entities themselves, we typically represent the entities by names which can be conveniently stored and manipulated.

Events that occur in a data base system are, therefore, models of events in the real world.

Since a data base is a model of the real world, in order for the model to remain current every real world event must result in a corresponding data base event. That is, if and only if a new car is added to our inventory, is data about the new car added to the data base. In fact, no matter what applications, user-interfaces, or data base management systems exist between the real world and the stored data base, this correspondence must continue to exist.

Thus, at the presentation-independent level the contents of the data base and the data base activity are already determined and integrated.

If we apply the new data base technology to viewing an organization's information <u>independent of its representation</u>, we find two important revelations:

(a) The information that the organization is currently using to support its business operations, regardless of its form, is the information it needs to achieve its present level of success. And even if the existing implementation of that information contains inefficicies, and redundancies, correcting these via data base technology will not alter the substantive needs of the organization.

(b) Because the only characteristics that separate existing files are those of representation, at the representation-independent level, an organization's information already appears as a single, integral collection. Furthermore, by factoring out the disintegrating characteristics, we gain a better perspective on the importance or worth of integrating the representations.

The migration to an "integrated data base" environment is consequently a task of improving the representation of information and the mechanisms for mapping to the representation from the real world.

If we model both the organization's information as well as its existing information management techniques, we will find that no matter how things are being done, there is some form of representation and there is some mechanism for mapping real information needs/updates into it. By isolating these and analyzing where they are bottlenecked, we can very systematically plan and implement improvements. And these improvements need not necessarily be implemented all at once or via a DBMS. They could be as simple as combining two redundant tape files and making the appropriate program modifications. Where the advantage of a DBMS comes in is that:

(a) To the degree that the DBMS insulates the application programs from representation changes, the program modification costs we incur as a consequence of improving the representation will decrease accordingly.

(b) To the degree that the DBMS provides latitude in the choice of representations to which it can map, the direct costs of improving the representation will decrease accordingly.

In this sense, the DBMS is only a tool that aids the task. There are many other such tools and we have great need for many more.

Question 3. How can I effect a shift within the management of the organization into a data base oriented environment? What are the sociological implications? What logical steps can I take to educate users?

In group discussion of this agenda item, it was clear that over the past several years there has emerged a growing level of consciousness of the role of information, and of its effect upon the sociology of the organization of which it is a part.

The only effective approach to a data base environment is to effectively sell management at the highest level. While some top level management would be interested in the complexities of the software, this area should be minimized. Normally, the results-oriented executive would be sold on the basis of providing answers to some of his periodic requests in a relatively short order, coupled with a carefully compiled benefit analysis. Middle management should then be invited to participate in a presentation and discussion on how DBMS can assist their operating units. This participation is necessary for a convincing sale.

There was a generally accepted level of understanding over the use of information as a power base in an information-oriented society and organization:

The use of data base technology means a major change in operational procedures to most organizations. Where this is the case, a variety of sociological factors must be dealt with.

1. A natural resistance to scrapping old methods in favor of such a radical new methodology.

2. A resistance to the loss of ownership of data which is implied by data integration.

3. Allied with the loss of ownership of data is the loss of total control over its contents.

4. Systems designers will find their job much easier. They will not need to, nor will they be able to, design the data base.

5. Programming will be easier, implying that a smaller programming staff will be needed.

6. The answer to questions will be much more readily available through the use of the query language. This will make it possible for users/customers to bypass the programmers and systems designers in many cases, thereby giving the users greater control of some aspects of data processing than they have had in the past.

Education of those involved in the implementation of more effectively managed data base environments was discussed and felt to be sadly lacking.

1. Top management must make known their firm commitment to the new technology, and must reassure users that they will not be adversely affected by the change.

2. Presentations must be made in the form of lectures and seminars to acquaint users with the new technology and its potential advantages to the users.

3. The users must be actively involved in all subsequent decisions regarding the design of the data base system.

Technological training is needed in data base design and its systems design implications. Education about the future directions that systems are likely to take, such as distributed processing, would be very helpful. Some college courses are now being offered, but most available training right now is in the form of seminars given by private firms.

Sociological training would be helpful in assisting the manager to deal with such profound change. This training is available both through college courses and private firms.

Question 4. What about data base administration? When should I create such a staff? What is its role?

There was a general agreement that a role existed for some concentrated involvement of a full-time staff to be the central manager of the data base, but specific duties were still illdefined. To a large degree, the role of data base administrator depends upon the tools applied to the task:

The data base administration staff should be created just as soon as the decision is made to use data base technology. The data base administration staff is responsible for the organization's data. It is responsible for the design of the data base, the integration of the data elements, the contents and use of the data dictionary, and the documentation of the contents and structures of the data base. It is responsible for ensuring adequate backup and recovery procedures and for controlling and maintaining the passwords. It ensures that adequate accuracy controls are present in the systems design.

Some very specific functions were described, clearly the result of some harsh experience:

Data base administration is required by a functional change within the data processing environment brought about by the transition from batch oriented environment with basic single user files to pooled files and multiple users. The data base administration is responsible directly for:

(a) Integrity of the data
(b) Data base backup
(c) Data dictionary
(d) The addition, maintenance, and deletion of the data elements
(e) Structure

The function of data base administration should be created as part of the decision to enter the DBMS environment. The staff should be kept as small as possible and composed of technically oriented personnel with the capability to successfully communicate with all levels of management. The function should report to upper management rather than data processing management.

Some disagreement existed on the need for the full-time DBA staff, clearly a function of the nature of the environment.

Assigning a person or group of persons to work exclusively in the role of data base administration depends on the motivation of the user for utilizing a data base management system. If a data base management system was installed to support centralized control of data, then a central authority (the DBA) for enforcing this control is necessary.

If, on the other hand, the motivation for selecting a data base management system is solely to acquire simplified methods for handling complex data structures then the role of data base administrator, as a full-time job, is not as essential.

Every user, however, who had implemented an explicit data base management environment, had established a role of data base administrator. The real significance of the role, consistent with the implicit benefits of data base technology, is the increased awareness of the central control authority for the resource known as information.

Question 5. Should I use a commercial software package? If so, what are relevant considerations? If not, what are my alternatives?

The obvious alternatives to using commercial software packages are expensive, time-consuming, and demanding of long-range commitments to maintenance. There was a general consensus that writing your own data base management system was an untenable solution unless the situation was absolutely unique. The dilemma was very eloquently stated:

Present data alternatives to data management are perplexing.

Practitioners faced with plotting a future course of action with respect to data base management have a formidable task. Commercial vendors now offer us nearly a hundred generalized data base systems, none of which has ever been demonstrated to be superior. Researchers tell us that the ultimate solution is at hand but, as yet, have produced only promises and prototypes. And on top of these, we had more than 10 years of historical success managing large, complicated, on-line data bases before we even knew there was such a thing as data base. Should we continue to use traditional file techniques, should we convert to an available DBMS, or should we wait to see what new alternatives appear?

Selecting from among candidate alternatives involves both qualitative and quantitative considerations.

Ideally, all considerations relative to selecting a data base alternative should be quantified, combined into a composite figure-of-merit, and applied to an objective decision. Eventually, this will be the case. Unfortunately, we as yet do not know how to quantify many of the considerations. And typical of most scientists, when we don't know how to quantify something we say it is qualitative. Presently in this category we lump criteria such as degree of separation, ease of use, privacy, integrity, and the entire gamut of categories typically seen in weighted scoring matrices. Even the hard, quantitative characteristics such as response-time, storage space, and CP/IO utilization have, in the past, proven difficult to predict and virtually impossible to compare.

A somewhat less eloquent, but equally incisive, viewpoint was expressed:

The concept as it stands today can be described as "muddled" leaving you with the impression that Mr. Eugene Pierre was correct when he wrote in the Honeywell Computer Journal, "The only thing standing between you and your successful Management Information System is your current management; your current information; and your current system." Indeed, determining the need for and choosing a generalized DBMS package is today delicate.

If the DB concept is just a collection of records; it's an old concept. If the concept is CRT inquiry and "magical" appearance of Management Information, it's an old concept. In fact, conceptually over half of what is today hawked as new data base concept has been around for years. We have had storage devices with I/O software to drive them; mainframes with OS's and support routines; programmers and programming languages; end users and reports, CRT's, TTX's, etc. What's new?

Some progress has been made in evaluation of the relative merits of data base systems by several research outfits. The one represented at the conference was Martin Marietta where a database simulation process has been successfully implemented. The simulation, however, cannot successfully compare those database characteristics which are the most subjective:

The qualitative characteristics, as yet, cannot be compared with total objectivity. These are immensely important considerations if we are trying to determine the cost-effectiveness among a set of systems. The mapping between these factors and ultimate expenditures, however, is difficult to construct. Take stability for example. To objectively compare two systems, we would have to hypothesize some changes to a real world information structure and measure, for each system, the amount of labor/computer resources necessary to modify the existing application programs. To do this in the general case, we would also need a measure for the difference between one information structure and another. These are not trivial tasks, but in light of continued dependence on the highly subjective scores and weights methodology, it is mandatory that we undertake them.

Question 6. What are implications of:

(a) changing and evolving standards;
(b) evolving hardware technology - mini-computers, mass storage, telecommunications;
(c) software evolution.

The consensus was that we, as practitioners, must continue to forge ahead with more effective data management techniques and processes, notwithstanding the erratic and often frustrating efforts of hardware and software developers:

Hardware advances will not change the data base problem-they will just relocate its solution.

Of all the advances in hardware that have pretended to be the end of the software industry, few have ever come to pass. And even if they are all eventually realized, data base technology will still be an important discipline. Consider, for example, the one advance that has been talked about as harboring the greatest impact on data base - associative memory. With today's memories, the whole data base issue arises from the fact that information structures are many-dimensional while our existing storage is essentially one-dimensional. Imposing a one-dimensional structure on one-dimensional storage is the principal problem that DBMS's attempt to solve. With associative memories, some number of associations greater than one will be provided by hardware but you can bet that the dimensionality of the information will always be greater than that. And even if all the associations can be accommodated, the techniques by which the hardware designers accomplish this feat will probably be very similar, indeed, to what we do in software today. The only possible impact I can visualize is that hardware may eventually become so fast that data base optimization will no longer be necessary.

The evolution in software technology will be dramatic. We must somehow ease the pains of migration.

In the next decade, there will appear some genuinely imaginative approaches to DBMS. It is possible that they will only be found in the puzzle palaces due to their unprofitability in light of DBTG, but they will, nevertheless, be there. At the representation-independent level, we have already seen several relational prototypes. As some of the new work in binary associations is completed, we can expect to see user-views with even greater stability yet. At the representation level, we will see DIAM-based systems that offer complete representation flexibility and search path optimization. We will see high level query languages used for host language interfaces and even for the description of data representations. We will see generalized data translators and hopefully, even time-shared/in place versions of these. The list goes on but one point becomes increasingly clear. With developments breaking at this pace, we will simply not be able to afford the conversion costs required to keep abreast of the latest technology. Yet with the proportion of our DP budgets that is devoted to data-related computing and data-related program modifications, we also can't afford not to.

We clearly require some interim solutions that will make these advances more accessible. Surprisingly, there are some rather obvious things in this direction we might do.

If we appeal again to the principle of representation-independence, there should be a way that advances in representation could be accommodated with little or no impact at all. If, for example, we selected one of many proposed representation-independent query languages and used it as the input to a generalized search path selection algorithm, we could, today, use any of the low-level procedural DBMS's (such as DBTG, TOTAL, IMS, etc) as interchangeable, "plug-to-plug" compatible packages (presuming, of course, we could perform the data translation process efficiently). Furthermore, any future changes to these systems would be completely invisible to the application programs. We could go one step farther and recognize that all the existing relational languages are so similar in structure that, if we used a relational language as our interface, we could also plug in any relational system that appears with virtually no modification (or none whatsoever with a minimal syntactic translator). Even if a system based on binaries appears, there exist some straight-forward n-ary to binary mappings that could suffice as a temporary interface until the applications could be modified. There are surely other possibilities than the ones I've suggested and some attempts in this direction are warranted.

The reaction of users to the efforts of the standards community was considerably more intense:

Standards produce a "wet blanket" effect in any industry.

Standards are the product of good intentions, but anytime they cross paths with innovation, the result is stagnation. Look at the building industry. Uniform building codes were very useful until the appearance of such innovations as foam houses, modular construction, individual sewage treatment plants, plastic plumbing, and solar heating. Regarding these, the UBS's, at best, got in the way and more typically drove the innovations out of business. Granted, some of the truly beneficial inventions did finally get a foothold, but their advent was delayed by years.

We seem to be in a similar position in database. The CODASYL people did an admirable job considering they did not have a crystal ball with which to predict the then imminent breakthroughs in conceptual technology. However, both the breakthroughs and the DBTG specifications are now behind us and the net effect on our organization and, I speculate, on other vendors as well, is that many aggressive DBMS development projects have been postponed or scrapped altogether. Our reasoning is that, by the time we could have a production version to market (say 2 years), the CODASYL DBTG inertia will be too great to overcome. How do you convince a DP executive, who has just spent two million dollars converting to CODASYL, that he ought to convert to something better?

Another major user expressed considerable concern over the efforts of the standardization community:

To me the Data Base concept involves all of these things and no single solution with any acronym will solve these problems to my satisfaction. I guess that what I'm saying is that, in my mind today, the elusive generalized DBMS is 60% concept; 10% software; and 30% marketing hype.

CODASYL is approaching this list of jobs to be done with an eye toward developing generalized language in each area.

AREA	LANGUAGE
Storage	Storage Structure (SSL) Device Media Control (DMCL)
Operating System	Control Language (OSCL) Logical
Logical Structure	Schema and Sub Schema (DDL)
Data Manipulation	Data Manipulation (DML)
End User	End User Facilities (EUFTG)
Administrator	Management Tools (DBAWG)

Remember that the CODASYL goal is to attempt to "generalize" language in these areas, not to "standardize". If the CODASYL solution works, gains recognition, is implemented and used widely, fine...if some other source provides a better solution, CODASYL wants that to become the standard. There is no future in "pride-of-authorship" selling within CODASYL. The products must stand on their own merits.

To understand the current standardization climate, I'd like to compare Data Base progress with COBOL progress.

DATA BASE	COBOL
1965 - Trying to understand problem	1959 - Already understand problem
1971 - DBTG report - no implementation	1960 - Published report

1973 - DDL Journal published - a few implementations

1974 - Standards Activity starts - a few implementations - no widespread use

1961 - Several implementations

1963 - Publication - many implementations

1965 - Publication - standards activity begins

1967 - Draft standard - many implementations

Today's situation is essentially primitive. The American National Standards Institute (ANSI) is gearing up to look at various products by identifying potential slots in which to conceptually place them. In short, they too are still trying to strictly define the DB problem.

The sense of the working group, however, was one of optimism and energy. Clearly, data base evolutionary activity has gathered momentum and proponency, and has begun - at least - to police its own activities enough to feel comfortable about calling itself a discipline. The essential thrust of the group was to gain greater recognition that more effective data base management was an evolutionary process and was not inevitably related to data base management systems, on-line access, or telecommunications. Progress along that evolutionary path had to continue in spite of advances which might eventually overtake some segments of ongoing activity.

3.4 Postcript.

The Chairman of the working group on User Experience is deeply indebted to the active participation of each of the members, and particularly wishes to acknowledge the following direct contributors to this report:

Lowell Schneider, Martin Marietta
Thomas Duff, Honeywell Information Systems
Ruth F. Dyke, U.S. Civil Service Commission
Roger J. Kelly, N.Y.C. Comptroller's Office
Richard Kurz, Southern Railway System
Margaret Derby, U.S. Civil Service Commission

4. STANDARDS: A DATABASE IMPERATIVE


STANDARDIZATION

4. STANDARDS: A DATA BASE IMPERATIVE

Working Panel Report on Standardization

Chairman: Robert W. Bemer

Biographical Sketch

Robert W. Bemer is a Senior Consulting Engineer with Honeywell Information System, Inc. His extensive list of accomplishments include:

Director of Programming Standards at IBM in 1962,
Developed original scope and program of work for ASA X3
  and the ISO TC97 standards body,
Chairman, TC97/SC5, Common Programming Languages.

In addition, Mr. Bemer was editor of the Honeywell Computer Journal and the publication, "Computers and Crisis." Earlier in his career, while at IBM, he developed COMTRAN, a predecessor of COBOL, and XTRAN, a predecessor of ALGOL. He is now chairman of the ANSI SPARC Study Group on Text Processing.

Participants*

Thomas Bergin	Chester Smith
R. E. Blasius	Lee Talbert
Milt Bryce	Alan Taylor, Recorder
Jeffery Ehrlich	Ewart Willey

4.1 Terms of Reference.

Because the working group was requested to project the status of DataBase System** standards in the next five years, the membership was formed of selected active experts who are familiar with past and present standardization efforts in the computer field. Moreover, the membership was deliberately selected to include international views and experience.

The forecasting requirement in the terms of reference required the group to consider the perceived need for successful and safe database usage. All agreed that there was every indication that the current increase in database usage would continue, and that this would be beneficial to commerce and government in all countries. Provided, however, that some way existed to ensure that the users of such databases could have confidence in the validity of information produced without having personally to undertake the impossible task of understanding all of the complexities involved in the creation and operation of the database, as well as the use of the data stored there.

---

* Complete addresses and affiliations are in Appendix C
** A neologism; see section 4.4.1.

Standards were seen as a method of providing users with such confidence. Accordingly, the working group focused upon the realistic and attainable standards that current technology could be expected to provide, in this time period, to promote and protect safe database usage. The need to anticipate still unknown technological developments (a need implicit in all standardization processes) was regarded as part and parcel of this task.

4.2 Basic Premises.

- Database standards embrace more than "management"

Database standardization activities are expected to cover all aspects of database usage, rather than just the narrow emphasis upon database management that has until now taken up most of the activity in the U.S. and other standardization groups. The already-developed CODASYL work on Data Description and Data Manipulation Languages offers a more-than-acceptable technical basis for standards. Because technical standards of some sort are prerequisite for any protective standards for database use, the working group believes that the perceived urgent needs for such protection will be based upon the CODASYL and related work.

- Database standards are an international concern and responsibility

The identity of problems across international borders, a basic corollary of the easily-perceived identity of computer benefits that have similarly passed from nation to nation, makes it both likely and advantageous that the standardization work should be coordinated from an international, rather than simply national level. The volunteer effort that has fueled national effort in the past will not be able to cope fully with the apparently inevitable trend to internationalize database standards. The urgency and economy of obtaining internationally-agreed standards should, and do, more than justify the small amount of new funding required for their development.

- The monetary and social aspect of database standards is large

It is difficult to calculate actual benefits of international protective standards, which can provide both safe operation of current databases and a safe, economic transition to the use of new hardware and software developments as they arrive, but we know them to be very great. Unprotected database usage has no real way of either assuring the integrity of the operation or protecting large investments in databases from being reduced or destroyed by technical obsolescence. Nor can we achieve the benefits from reducing training requirements, providing easy interchangeability, and using newer technologies that permit users to choose between central and distributive philosophies for database operations.

- Inhibitory effects of standards are small compared to benefit derived

Unprotected databases cannot satisfy economically the social demands for security and privacy that spring from the technological capability to make data independent of the media upon which it is stored, or independent of the organizations, structure, hardware, or software involved. These deficiencies, when placed alongside the benefits of protected database usage, make it clear that temporary or apparent restrictions implicit in the standardization process are minor in comparison with the benefits to be gained from standardization. Therefore the standardization work should not be delayed or inhibited by arguments of restrictiveness.

4.3 Organization of the Report.

The organization of the balance of this report is intended to lay out just what the working group believes should be standardized -- who should do the work, how it should most properly be undertaken, when it should be reasonable to expect standards to emerge, both formally and informally, and why the group believes that standardization should and will proceed in this manner. This format of presenting the report as a set of recommendations was chosen when it was realized that the nature of the current need for standards in this area permitted no weaker position to be taken, in view of the working group's professional obligations.

4.4 Standards Expected for Database Systems.

Four separate groups of standards were recognized as necessary to the development of database usage during the next five years. These are:

- Terminology Standards
- Criteria Standards
- Component Standards
- Usage Standards

As a general consideration, all database standards should avoid being tied to particular programming languages such as COBOL and Fortran. However, dialects peculiar to associations with such languages are considered harmful only when they result in the loss of essential benefits, such as the ability to transport a database for use by another host system via purely mechanical conversion of the dialect into pure standard form. This general consideration requires that the technical characteristics of host-free standards be observed.

4.4.1 Database System Terminology Standards. Terminology standards are urgently needed, not only to improve communications between active and prospective database users, but to facilitate understanding in the academic and development areas of database study. A single developer-oriented set of standard terminology is not considered to match these requirements, although the standard user terminology must necessarily be a subset of the larger set of terms used by developers and researchers. While terms such as "schema" are perhaps allowable in early development (although why this should be so is not clear), the fact that such concepts have to be understood by businessmen charged with day-to-day decisions from databases of all types makes it imperative that developers be prepared to find and accept less esoteric terms as soon as the concept becomes qualified for inclusion in the user standard. This need for better user communication should, where appropriate, even force withdrawal of terms previously accepted by developers, and replacement throughout both developmental and user communication by more commonly understood terms.

As examples, the working group makes two strong recommendations:

o The term "Data Base Management System" should be changed. Management is only one part of the proper subject, which includes development, use, interchange, protection, etc. of databases. The focusing of attention upon the technological controls utilized with data bases has almost hidden the scope of the effort necessary to permit the benefits arising from their use. The term "DataBase System," abbreviated to "DBS," should replace the old and inaccurate phrase in every terminology.

o Database systems not using computers have been in existence for millennia, and will continue to be used throughout the period under study. Therefore the terminology "CDBS," for "Computerized DataBase System" should be used to refer to the software components of such systems and all the other tools necessary to provide a DBS operating via computers. Moreover, wherever possible the characteristics of noncomputerized systems should be studied and referred to by standard terms that mean exactly the same as they do for computerized systems, and full precautions should be taken in the selection of terms for computerized systems to ensure that the processes of noncomputerized systems can be identically described and studied.

4.4.2 Database Criteria Standards. Criteria are needed for use in the evaluation of proposed database standards, particularly in view of the variety of interests involved, e.g., implementors, users, auditors, management, government regulators, etc. These can be developed ahead of the actual component standards; they can then be used to provide better understanding of, and better direction for, such standards and their development.

A model of such a Criteria Standard List is the document used by ISO/TC97/SC5, the international body for standardization of programming languages. It was considered of sufficient benefit to reproduce it in this report as Appendix A. It may also supply some modifiable text suitable for the database criteria document.

4.4.3 Component Standards. Database standards will not be monolithic. Separation into readily-identifiable parts is desirable for both data independence and flexibility to adapt to new developments. Some of these are:

- o Data Elements
- o Data Dictionaries
- o Data Description Language
- o Data Manipulation Language
- o Archiving Methods
- o Data Transmittal Methods (via portable media such as magnetic tapes, or via system-to-system communication methods)

4.4.4 Usage Standards. The protection of the integrity and other vital characteristics of databases involves many activities whose performance can be made practical and efficient only by having standards. This is because some of the needs of these activities become economically providable only if done in standard ways. These include:

- o Validation of conformity to technological standards
- o Auditing
- o Social implications (integrity, life cycle, accuracy, completeness, etc.)
- o Diagnostic procedures
- o Guidelines for proper usage and administration
- o Registration of common data structures
- o Structure convertability
- o Performance measurements

4.5 Actors and Activities in DataBase System Standardization.

The entire standardization activity in database systems will have the contributions of six classes of organizations:

- o Data Processing professional associations (ACM, DPMA, BCS, CIPS, with their supergroupings and federations such as AFIPS and IFIP, augmented by ad hoc groups such as this working group and this conference)
- o Professional associations with a strong data processing aspect (e.g., AICPA, IIA)
- o Developmental and user groups (CODASYL, SHARE, GUIDE, JUG, and the user groups of other computer manufacturers)
- o Groups specifically developing and approving computer standards (ECMA, ANSI X3 (CBEMA))
- o National governmental bodies (U.S. National Bureau of Standards, the UK's CCA, Canada's CGESC)
- o International governmental bodies, arms of the respective departments of state (CCITT, ICA, etc.)

The group concluded that while the ISO is the registry and approval body for international standards, it supports no developmental effort to produce the candidates for standards. Because of previous conclusions that rapid development at an international level was most desirable, the ideal situation pointed to the ICA, the Intergovernmental Council on ADP, which has a charter directly related to databases. In view of the influence and viability of the CCITT in worldwide communication standards, it was felt that the ICA could take the lead role, with the help of a permanently-funded organization. It would rely for professional direction, but not for development funding, upon the existing volunteer operations where standardization has been considered until this time.

Under this scenario:

o The national standardizing bodies, and ISO, will continue in their formal role.

o Professional bodies will contribute more tutorial papers, hopefully covering a higher ratio of social to technological aspects than has been the case until now. Tutorials are a vital complement to any formal standard; they enhance acceptance, minimize confusion, and sometimes show possible improvements for the standard.

o International free-standing bodies in the CODASYL style, and particularly CODASYL itself, will continue to develop much of the content that will later become the bases of national and international standards.

o Governmental bodies having specific charters to assist administrative and executive agencies with technological planning will take stronger roles in speeding test usage of standards proposals prior to formal adoption procedures, and in promoting wider usage subsequent to such adoption. NBS is a particularly valuable and active example, and we recommend and expect that its activity here will be increased and expanded.

o ICA should support the management and control activities necessary to move the database system standardization activities through all the necessary processes and steps at maximum practical speed.

4.6 Expected DataBase System Standards.

If any database system standard is to be created and adopted during the 5-year period, it must be based upon the output of the CODASYL group, already available. No other candidate has been formally entered into the standardization procedures. To the knowledge of the working group, almost all other systems in use today are proprietary. Here we note a sharp parallel to COBOL, the original effort of CODASYL, where the proprietary packages were displaced completely because COBOL was a "standard," where the others were not.

Another reason why the CODASYL work is the only candidate comes from experience in international standards for programming languages, which indicates that 8-10 years of usage is a practical prerequisite of any basis for an international standard. The CODASYL work meets this condition.

Therefore, to get an approved international standard in the time frame that the working group feels is vital, that is, by mid-1977, the only effective basis is the CODASYL work. Another necessary condition is that the NBS and other appropriate bodies signal their support of the activity very soon. The proposal must be taken to the ICA for its consideration.

4.7 Modus Operandi - How to Get a Standard.

The working group agreed:

o Public interest in two aspects of database usage--privacy and governmental/business decisionmaking--was so high that standardization in this area was both vital and urgent.

o It is clearly a matter of international importance (both the UK and Canada were represented in the working group).

o Therefore the voluntary, intermittent working of individuals contributed by their employers is not a viable method of achieving the necessary standard in the necessary time frame. A permanent directorate is needed, to plan and schedule the multiple actions necessary to meet the goals. The effort should be international, properly coordinated, and closely monitored.

The conclusion on modus operandi was:

o An international body is required as sponsor and/or directorate for database work. Here it is noted that the CCITT receives government funding support for similar work in worldwide communication standards, being an arm of the several departments of state.

o The ICA (Intergovernmental Council on ADP) was acknowledged to be a body that could act analogously to the CCITT. Such authority should eventually reside here. However, preparing to do so could take a period of time constituting an unacceptable delay.

o It is therefore recommended that the NBS, as caretaker for the ICA, undertake the organization and coordination functions required. The NBS has provision for Research Fellowships, enabling continuous attention to the project in all aspects--tutorial, scheduling, terminology, criteria, and coordination with standards bodies. The basis of the work is to be the output of CODASYL and the just-inactive ANSI SPARC group.

o It should be recognized that the U.S. Congress has already appropriated funds for standards in computer usage, and that there are commissions appointed on privacy and electronic funds transfer, both of which relate almost entirely to databases.

4.8 Supportive Arguments.

The following are given in support of our position:

- o There is no clear indication that multiple standards are either necessary or desirable in database systems. We wish to avoid at all costs the partial duplication suffered in the standardization of programming languages.
- o In the absence of definitive standards and descriptive material, users are subject to pressure to use diverse proprietary packages. False starts have been many, with expensive conversion and restructuring.
- o Investments in database design and usage are even greater than in programming language applications.
- o The training requirements for database operation are substantial, perhaps even costlier than the computer investments. Arbitrary and capricious differences are confusing and costly.
- o The necessary body of knowledge for data independence exists. Logical structure must be divorced from physical structure for reasons of transportability and future architectures. The users can be implicitly protected for this by a standard.
- o A standard will insulate from many costly dangers.
- o Specifying a host-free standard also protects investments.
- o Controlling diversity in database system usage will also control diversity in auditing and other control procedures required, thus enabling concentration on excellence, not diversity, for limited resources.

4.9 List of Active Working Groups in DBMS Standards.

ISO/TC97/SC5 Working Group on DBMS:

Secretariat - US, c/o
Marie Hogsett
ANSI
1430 Broadway
New York, New York 10018

French Working Group Z 6/SC5
AFNOR
Tour EUROPE
CEDE 7
92080 Paris La Defense
FRANCE

BSI-DPE Working Party on DBMS
British Standards Institution
Maylands Avenue
Hemel Hempstead
Herts. HP2 4SQ ENGLAND

ECMA Database Committee TC22 (ex TGDB of TC6 COBOL):

ECMA
L. Lauri, Technical Officer
114 Rue du Rhone
1204 Geneva
SWITZERLAND

ANSI/SPARC/Data Base Study Committee (in suspended animation)

ANSI/X3J4 COBOL:

ANSI/X3J4
c/o Robert Brown
Director of Standards
CBEMA
1828 L Street, N.W.
Washington, D.C. 20036


AUDIT

5. AUDITING THE DATA BASE

Working Panel Report on Auditing

Chairman: Donald L. Adams

Biographical Sketch

Donald L. Adams is Managing Director of Administrative Services at the American Institute of Certified Public Accountants, with responsibility for internal applications of the computer as well as development of its use in the accounting and auditing practices of members.

Before joining AICPA, Mr. Adams was assistant director of data processing at the investment banking firm, Salomon Brothers. Prior to that, manager of computer auditing at Peat, Marwick, Mitchell & Co. His interest in computer auditing spans sixteen years. He has written many articles on the subject, has lectured extensively in the United States, Canada and Europe, and edits the monthly newsletter, EDPACS (EDP Audit Control & Security).

Participants*

Peter M. Benson	Don Lundberg
Adolph F. Cecula, Jr.	John Nuxall
Dennis Fife	Robert Stone
Tom Fitzgerald	Ian D. Watson
Dick Hirschfield	Ron Weber
Ted Hollander	Harold Weiss, Recorder
Albert A. Koch	Luc van Zutphen

5.1 Background.

The people who participated in the workshop on audit considerations spent two days and a portion of one night developing a consensus about the major impact of the data base on the auditor. In developing a report based upon these deliberations, the final format was left to the discretion of the Chairman.

Fully exercising that discretion, a format was created. This report will be presented in three basic segments:

A brief article outlining, in very broad terms, the workshop Chairman's view of the auditor's concerns in regard to the data base. This article is solely a representation of the Chairman, and does not necessarily reflect the views of anyone else who participated in the workshop.

---

* Complete addresses and affiliations are in Appendix C

A summary of the consensus was developed during the workshop. Six basic areas were addressed:

1. Differences - Data Base Systems vs. Others
2. Integrity Considerations
3. Audit (Management) Trail
4. The Auditor's Role
5. Interface with Audit Software
6. The Next Five Years

Selected questions and answers were prepared prior to the symposium. In order to start the ball rolling prior to the workshop session, the chairman prepared a list of questions for consideration and a number of the participants submitted their answers to those queries. A summary of the most interesting questions and answers is included as the final part of this workshop's paper.

5.2 Audit Concerns in Re Data Base Systems - An Overview by the Panel Chairman. Auditors have always been concerned about control, security, and integrity. Would the presence of a DBMS have a special impact on their efforts or obligations? The foregoing is a reasonable enough question, but one that is sometimes difficult to answer. Many auditors who have been exposed to data base systems feel they do have a significant impact on the audit process, but it is hard to articulate sound reasons in support of this feeling. In order to establish a common framework for consideration of these potential audit problems, it may prove useful to categorize them as follows:

5.2.1 Security.

5.2.1.1 Access. In a DBMS, all the information eggs are in one basket. This is a basic element in such systems. One of the big selling points is the fact that anything anyone wants to know is in one place. This has been called the "Corporate Data Bank" approach. In spite of its positive aspects, this feature is also a drawback. A fundamental element of internal control is separation of duties. This concept is equivalent to the "need to know" basis applied to military security. A data base, by its very nature, does not contribute to the separation or compartmentalization of data. Therefore, it tends to weaken control.

If an individual has access to all data in support of an organization's activities, he will find it easier to manipulate those records to further his own purposes. Now, this gets into the area of fraud. External auditors are not responsible for detecting fraud. This is quite clear. However, when the potential for fraud is greater, the auditor should at least be aware of this fact and consider it as a factor that should influence the study and evaluation of controls. The auditor should review access controls and evaluate their contribution to the separation of duties.

5.2.1.2 Update. Users who are authorized to update the data base intensify the problem of a reduced separation of duties. The potential for error, whether deliberate or not, is much greater. The auditor should conduct a detailed review and evaluation of control over updates.

5.2.2 Integrity. Two aspects of integrity should be considered:

5.2.2.1 Input Errors. Auditors have always been concerned with the accuracy of data, so a review of input controls is fairly standard. This also involves a look at controls over errors. If a transaction was in error, it should be corrected and put back into the system. A DBMS does not change these concerns, but it does complicate the processing that is involved.

Input Control. In non-DBMS applications, the same transaction often serves as input, in different formats, to a number of files. During the normal course of processing, whenever two files contained the same data fields, there was chance to compare the two and detect errors. Similarly, if an error was suspected, it was often possible to compare similar files and isolate those records that might be sources of the error. Neither of these abilities is available within a DBMS. Control over initial input must be better in a DBMS or the overall accuracy of the data will be reduced. The auditor should be aware of this possibility when reviewing and evaluating input controls.

Error Control. In non-DBMS applications, transaction files were often processed against master files and, as part of such processing, errors would be flagged. However, the item in error was still on the transaction file and was still part of the control totals, so it had to be accounted for within the system. DBMS processing is often designed to prevent an error from being recorded on the data base. This can lead to the loss of error transactions. The auditor should be aware of this potential problem.

5.2.2.2 File Integrity. Several years ago, Harold Weiss wrote an article in which he coined the phrase, "total corporate amnesia." He predicted that, some day, a large company, wedded to the fully integrated system approach, would lose its data base, not be able to recover, and would go out of business because it no longer had any data. Logically, this will happen some day. There have already been a couple of near misses.

DBMS are quite fragile. When a complete disaster does occur, it will probably be caused by a systems programmer misapplying a modification to the DBMS software or making a mistake during a revision of the data base structure. It might be caused by an operator during an attempt to recover from a data base failure. At such times, the system is particularly vulnerable. A number of disaster scenarios could be constructed, but the overall measage seems clear.

The first time there is a case of a corporation with fatal amnesia, there will be an awful lot of finger-pointing when it comes time to establish who was at fault. There is sure to be a rash of law suits and the auditors, both internal and external, are bound to be named. If the auditors go to court, they will probably be saddled with part of the blame. A lot of people may feel that is far-fetched, but the plaintiff's lawyers can make out one hell of a case. For example:

(1) The auditor is deemed to be an expert in the workings of accounting systems.

(2) Audit work must include a review of the controls within the accounting system.

(3) Generally accepted accounting procedures require the use of the "going concern" concept in preparing financial statements.

(4) Without its records, the corporation was not a "going concern."

(5) Based on both expert knowledge and a review of the system, the auditor should have known the corporation could suffer total amnesia.

(6) The auditor did not take a "going concern" exception in the certificate.

(7) Therefore, the auditor either did a poor job or deliberately withheld important information.

(8) Guilty!

At this point, two counter-arguments emerge. One is, the same thing could have happened without a data base. The other is, the chance of such a disaster is so remote that the auditor would not feel obligated to report it. While the first argument has some validity, it is a matter of degree. Non-DBMS systems tend to have a large amount of redundant data. Copies of this data are likely to be geographically separated. With DBMS, there is a much greater concentration of data. The odds were against a disaster that could wipe out all of a corporation's records in a non-DBMS situation. It was possible, but only remotely. If it did happen, it was likely to be associated with a natural disaster and auditors are not accountable for acts of God, yet. In a DBMS environment, the loss can usually be traced to a human failing, and auditors are more responsible in that area.

Arguing that a DBMS disaster is a remote possibility, so the auditor does not have to consider it, is not particularly valid. Suppose a thousand companies kept all their records in offices located in a banana-republic. One day, a paper-hating general seizes power.

For no particular reason, he announces that once a year, on liberation day, the name of one company will be selected at random and all of its books and records will be burned in a giant bonfire to honor the revolution. A lot of companies would quickly move their records somewhere else or start making provisions for back-up. Now, the auditor for one of these companies would want to look at what they had done to protect themselves. If they had done little or nothing, the auditor would be gravely concerned. If this concern did not manifest itself in a going concern exception, it would at least rate a footnote. In any event, even though it is a 1000 to 1 shot, the auditor would not ignore the situation. Similarly, it is difficult to ignore a DBMS that does not have effective safeguards to protect its integrity.

5.2.3 Conclusion. Use of a DBMS intensifies the need to provide security and integrity features within the accounting system. The auditor must be aware of the increased hazards, and should consider them as part of the normal review and evaluation of accounting control. It is up to the auditor to decide what action to take in any given situation, but the potential problems associated with DBMS must be considered.

5.3 Audit Concerns in Re Data Base Systems-Consensus of the Workshop Participants.

5.3.1 Objectives. In attempting to address their specific charge within the framework of the symposium, the members of the auditing workshop decided they would direct their report to the manager, either EDP or non-EDP, who is considering the implementation of a data base management system. Basically, the members of the workshop planned to tell that manager what they see as the major differences between DBMS and conventional systems. Then, they would like to inform the manager about the integrity and audit trail features that are important in order to protect the organization using the data base and to provide for the needs and requirements of internal and external auditors.

To provide the manager with some additional perspective, members of the workshop decided it would be useful to present an indication of the auditor's role in the implementation of DBMS as well as an outline of some of the specific problems the auditor will encounter in working with a data base. Finally, the workshop's consensus output will conclude with a projection of developments that may take place in the next five years and what will impact areas of concern to the auditor.

5.3.2 Consensus.

5.3.2.1 Differences-Data Base Systems vs. Others.

Accountability. In traditional systems, particular files were associated with specific programs or applications. As a result, it was fairly easy to identify the person or group who was responsible for the input and maintenance of each data field. Because of the integration that takes place within a DBMS, it may be difficult to determine who is responsible for what. Prior to the implementation of the data, each data element must be specifically assigned to a person or group who will be responsible for its content.

Cross-impact. When systems were developed on an individual basis, there was a minimum amount of interaction between application systems. The implementation of a common data base creates a focal point in that all applications now interact with the same set of data. As a result, an application can have an effect, often a major one, on other applications. This increased integration makes the thorough testing of systems even more important than it has been in the past.

Increased Loss Risk. Full implementation of a data base will reduce or eliminate the maintenance of redundant information. The decrease in the duplication of data increases the opportunity for the total loss of that data.

Asset Value. Use of the data base concept, as has been pointed out, concentrates and integrates all of the information that will be used to support the decision-making process. This means that the data base is a very important resource within the organization. As such, it should be considered a valuable asset and be subjected to strict provisions for security and control.

Responsibility. Earlier, mention was made of the fact that the integration inherent in a data base may tend to obscure the responsibility for the data elements. Along these same lines, each application group will probably have less overall need to provide control over data, but must exercise a much higher degree of control over the specific elements for which it will be held responsible.

Control. The integrated nature of a DBMS will decrease the reliance on segregation or separation of duties as an effective control technique.

Organization. Support of the implementation and maintenance of a DBMS may require major changes in the structure of an organization, both inside and outside of the EDP department. New positions, such as data base administrator and user coordinator, may have to be created.

Data Definition. The use of a data base will improve the control over the definition of data since all such definitions will have to be standardized and coordinated. Further, because all access to data can be integrated, a DBMS provides better access control than that available in more conventional systems.

Volume Reduction. Elimination or reduction of data redundancy reduces the volume and variety of input to be processed and thus reduces the amount of audit time that must be devoted to the review and testing of input edit controls.

Audit Independence. In typical file organizations, the physical structure of the data must agree with the logical structure. In a DBMS, the two structures are completely independent. As a result, an auditor who wants to access information on a data base will be forced to use access methods supplied by the vendor who developed the DBMS. To some extent, this will reduce the level of the auditor's independence. However, the auditor would not be able to justify the cost and effort required to develop independent access methods.

Hardcopy Impact. Inevitably, the implementation of a data base will reduce the amount and extent of hardcopy audit trail that will be provided. However, this may well be offset by the increased use of online, machine-readable logs.

Stability. Because a data base cuts across all applications, it will be extremely difficult to implement major change in its structure without causing severe disruption of all processing. Since major change will be largely precluded, there will be a trend to more stable environments. This should help to improve overall control. At the same time, a DBMS can accommodate minor format or logical content changes much more easily than other systems can. This will make it somewhat easier to add or improve control features in existing systems.

5.3.2.2 Integrity Considerations. In reviewing the integrity aspects of DBMS, the members of the workshop focused on six major areas:

- Recovery
- Backup
- Quality
- Accuracy
- Controls
- Security

Recovery. The more data is shared between two or more applications, the more difficult and complex the recovery process becomes. However, difficult or not, a DBMS must have the capability of recovering from minor, or major, failures. This recovery only applies to failures caused by hardware or systems software. The DBMS cannot provide for recovery from failures caused by faults in the logic of application programs.

When the system is being operated in batch mode, the DBMS should provide two degrees of recovery. It should be able to back out all processing that took place since the last previous synch point, this feature can be used to handle minor failures, and, to cope with major failures, it should be able to back out the entire process.

Failures in on-line mode are more complex. Recovery is based upon a "unit of work" (e.g., program, block of records, a single record). In each case, either a unit of work must be defined to the DBMS or it must have the ability to impose such a unit on the applications being processed. Ideally, on-line recovery should be on a dynamic basis. That is, the system should be able to detect and recover from failure without the need for any outside intervention or assistance. Unfortunately, this ideal is not always possible. A failure that has gone undetected for a period of time may, because of the integrated nature of DBMS processing, have been compounded and propagated by the further processing that is based upon the original error. This kind of "associative failure" precludes the use of dynamic recovery techniques.

Hardware failure can involve either the storage device(s) used to record the DBMS or the execution of the software used for the processing. In order to recover from a storage failure, the DBMS should be able to reprocess a backup file forward until the point of failure has been reached or it should be able to back out processing from the failure point back to the most recent synch point. Recovery from a processing failure is much more complex. It involves the following steps:

- o Identify all transactions currently in process (in-flight transactions)
- o Back-out all completed processing that took place since the last synch point
- o Retrieve all output that was generated since the last synch point
- o Restart all processing

Back-Up. While the need for back-up is quite obvious, several key factors must be considered in developing a back-up plan. The frequency of back-up must be determined. Both the importance of the DBMS and its normal update or processing cycle must be evaluated in deciding on frequency. Monthly, weekly, and daily backup are common, but may not be appropriate in all circumstances.

Redundancy in back-up must also be considered. In a number of cases, two sets of back-up may be provided. One set will be kept on hand within the computer installation and the other stored at an off-site location. Log tapes, which are normally produced as a byproduct of DBMS processing, are an important part of any back-up plan. Since the log represents a record of all transactions processed, it provides the link between the back-up copy of a file and its current status. The logs can be quite lengthy, so it may be useful to utilize a DBMS utility that can be used to summarize and merge the details of the log while still maintaining its usefullness in providing for recovery. Logs must be retained for a reasonable period of time and it may be advisable to provide a dual-logging facility to insure redundancy in this phase of back-up.

When the DBMS software has been revised by its supplier, the revision will often affect the contents or format of the log. As a result, logs prepared prior to the most recent revision of the DBMS may no longer be usable for back-up purposes. Whenever the DBMS software is modified, all back-up facilities, particularly the logs, should be subject to careful review and testing.

Quality. In planning for a DBMS, the quality of data is a critical factor. Since data will only be recorded once and will then serve as the input for all subsequent processing, it is essential to maintain a high level of quality for all elements within the DBMS. Input data validation must be the initial step in any DBMS application. The editing should be as extensive as possible. Fortunately, since all information is available in the data base, it is easier to perform tests that involve correlation between various data elements (e.g., Is this an active account? Is the price for this item in line with the prior price?).

To some extent, the use of a DBMS will provide for techniques that will improve quality. Standard definitions of all data and the use of common validation or edit routines are virtually mandatory in a DBMS and both of these factors will contribute to improving the quality of data. As part of the effort to improve editing and quality control, a DBMS approach may cause a higher number of transactions to be rejected because they contain errors. Strict control over and follow-up on errors must be provided.

Accuracy. To the maximum possible extent, the DBMS should have a self-diagnosing capability. That is, the system should be able to detect and report on any deterioration of the data base (e.g., broken chains, scrambled pointers, or other errors internal to the data base). As part of routine operations, such as reorganization or the provision of back-up, the DBMS should be able to test and evaluate its own internal accuracy.

Controls. Certain control elements, which may be present in an EDP system, are particularly important in a DBMS operation. These include:

o An early consideration, during the initial stages of design, of the controls to be incorporated within the application system and within the data base. These controls must be designed to interface with other systems or applications, both now and in the foreseeable future.

o Such controls should be designed in accordance with a set of formalized control standards that have been developed for the organization.

o Normal accounting controls should be maintained in a data base environment. For example, provision should be made for reconciling data base contents with independently maintained control totals.

- Strict control over the authorization to perform critical data base functions (e.g., add, modify, delete) should be maintained.

- The system should, to the largest possible extent, be both self-controlling and self-correcting.

Security. Some special security considerations must be included in the planning and implementation of a DBMS. They include:

- All security violations detected by the system should be logged, reported, and investigated.

- A log that records and identifies all data base accesses should be maintained. While this may seem to be a burdensome requirement, it may become necessary in order to comply with legislated privacy requirements.

- Because no one individual or group can be considered to be the "owner" of the data base, responsibility for each data element must be established. In each case, a person or group should be given sole authority to grant access and update capability as it relates to a particular data element. From that point on, the person or group is considered to be the owner of that data.

- A current log of all access and update authorizations that are in effect should be maintained.

- The data base administrator, while responsible for the existence of the data base, should not generally be granted access to its content. In unusual circumstances, when such access is required, authority should be granted to the data base administrator in exactly the same manner it would be extended to any other user.

5.3.2.3 Audit (Management) Trail. In the design of a DBMS, early consideration should be given to providing an adequate audit trail for all processing that will take place within the system. As a user of the system, the auditor has the right, subject to normal cost justification requirements, to request specific reports and/or the creation and retention of files for audit purposes. Thus, the auditor may establish requirements that will result in the maintenance of an audit trail. However, the auditor must adapt these requirements to the economic realities of the system being audited.

Several key factors should be considered during the design of the audit trail. They include:

o A usable, complete record of all transactions that affect an account balance, or the contents of a master file, should be maintained.

o A balancing or summarization function should be provided as part of the system that maintains the audit trail.

o While the operating system and the DBMS both maintain logs, neither log was designed to provide an audit trail. Rather, they were specifically constucted to provide a restart and recovery capability within the framework of the system software. As a result, consideration should be given to creating and maintaining a transaction log specially designed to provide an audit trail.

o It is particularly important to provide an audit trail that can be used to control and follow-up on errors, rejected transactions and/or data, and items in suspense.

o Since the data base is particularly vulnerable when it is being reorganized, special audit trail provisions should be included in the planning and design of all such processing.

5.3.2.4 The Auditor's Role. Both internal and external auditors will become more deeply involved in DBMS, and both have a definite role to play in the design and implementation of such systems. The internal auditor is likely to become involved in the very early stages of data base design. This involvement will probably become quite deep, and a number of people will maintain that the auditor's independence has been impaired. To some extent, this may be true, but it is unavoidable. Without close and early involvement, the internal auditor will not be able to understand the system and discharge his responsibilities to management.

The external auditor will be involved as a user of the system and an evaluator of its controls. He will be using the system to provide input to the audit process. The review and evaluation functions are intended to provide the auditor with information that will influence the reliance to be placed on controls in determining audit scope. Further, the auditor will develop comments for presentation to management in regard to any weaknesses in the overall control scheme.

As was pointed out earlier in this paper, the basic roles of internal and external auditors do not change when a DBMS is implemented. Rather, there may be a shifting of emphasis within the range of functions the auditor performs. Some controls and operating procedures may become more important. Some audit techniques may become more complex. In any event, both internal and external auditors will require more technical training to equip them to perform an audit of a data base system.

Some areas of audit involvement or special effort in DBMS will include:

- o Keep informed in regard to proposed new systems
- o Evaluate audit, control, and security features in DBMS software
- o Perform or evaluate the cost/benefit analysis prepared to justify a proposed data base system
- o Test backup and recovery features incorporated in DBMS software
- o Review and evaluate minimum control standards
- o Determine the adequacy of data retention in regard to both management and audit requirements
- o Review the use of logs to see if it is effective
- o Test control exercised over system changes

5.3.2.5 Interface of DBMS with Audit Software. One real problem that auditors have had to face in regard to DBMS is the fact that the arsenal of computer audit software that has been developed over the years cannot, for the most part, cope with a data base file organization. Quite simply, computer audit software cannot read a data base. Some of the available packages do provide a data base interface, but, the use of this feature requires a higher level of technical expertise than that required to use the basic package.

To cope with their existing interface problem, auditors have utilized the following courses of action:

(a) For small data bases with a simple logical structure, use a utility or specially written program to dump the data base to tape in a sequential format and then use computer audit software to process the tape.

(b) Use vendor-supplied utility, retrieval, or report generators to produce information or perform processing for audit purposes.

(c) Develop, within the audit team, the technical expertise required to deal directly with a DBMS.

(d) As part of the design of the data base application system, build audit functions into the system. This approach is not, by any stretch of the imagination, in general use, but has proven to be quite successful in a small number of applications.

Each approach has advantages and disadvantages. From the standpoint of maintaining audit independence, (a) or (c) is the best choice. In regard to low cost and ease of use, (b) is probably superior. From the theoretical standpoint, (d) may prove the most promising in the future.

5.3.2.6 The Next Five Years. Being, by reputation and nature, a relatively conservative (in the non-political sense) group, auditors are not too comfortable with five year predictions. However, the group does see some likely developments:

o The dual growth of DBMS and public concern for individual privacy will create a flood of legislation. Someone will be given the job of determining whether or not specific data base applications comply with privacy regulations. Although they are not eager to accept this task, it seems likely that auditors will be called on to conduct such compliance reviews.

o Further development of fast, low-cost, almost infinite capity storage will make DBMS more practical and attractive. Many of the current problems of audit trail and the maintenance of historical files will vanish since everything will be kept online within the data base for a much longer period of time.

o The use of audit functions built-in to DBMS applications will increase.

o The standardization of data base software structure will eliminate most, if not all, of the audit software interface problems.

o Audit specialists, in much larger numbers, will develop the expertise required to work with data base software and applications.

Summary. If there was one thing the group agreed upon, it was that two days was not enough time to deal effectively with all of the audit concerns associated with DBMS. However, every effort was made to devote time to the most important issues and develop the consensus of thinking in regard to those issues. Hopefully, these efforts have produced information and commentary that will be useful to both management and auditors.

5.4 Selected Questions and Answers.

Question 1: How can we interface existing computer audit software with data base systems?

Viewpoint: First, it is technically feasible for developers of software packages to prepare routines that can access data directly and thereby completely bypass the DBMS. Although technically feasible, the cost to develop the data base access routines for each DBMS to be accessed would probably require complex consideration of both non-standard access methods used for disk, as well as assembly of various data elements whose physical storage usually does not have logical meaning until it has been processed by the DBMS. (If written, the audit software data base access routines would themselves effectively be functioning as a DBMS.) Although technically feasible, this approach may be impractical both from cost considerations and the level of expertise which would be required to use them.

Alternatively, the DBMS may be interfaced with the audit software for the purpose of extracting complete units of data (e.g., files) that can be totaled by the audit software package and reconciled by the auditor to independent sources such as general ledger control totals. Requirements for the design of interfaces will vary depending both upon the design of the DBMS and the audit software packages. Although developers of audit software packages may have to design their own interface, cost considerations clearly indicate the desirability of DBMS vendors providing a standard interface so that audit software developers would be required to make only minimal modifications to their software.

Viewpoint: Presented in Outline Form:

<u>MODIFY EXISTING AUDIT SOFTWARE</u>

<u>Advantages</u>

- o Familiarity of software to the auditor
- o Independence of the auditor is maintained
- o Interface is efficient

<u>Disadvantages</u>

- o Self-contained versus host language dichotomy
- o Lack of standardization of DBMS
- o Incompatibility of audit software language syntax with the semantics of the data structure models
- o Data definition used by the DBMS may be inadequate for audit software purposes
- o Independence is still compromised if the operating system access routines are used.
- o Integrity function of the Database Manager may be by-passed
- o DBMS environment is simulated so that integrity features such as concurrency control cannot be checked

<u>EXTRACT A SEQUENTIAL FILE</u>

<u>Advantages</u>

- o Familiarity of the software to the auditor
- o Independence of the auditor is partially maintained
- o Auditor is not responsible for the interface

<u>Disadvantages</u>

- o Integrity of the sequential file can be questioned

- Processing inefficiencies exist with the interface because indexes cannot be utilized, and sorts must be performed rather than pointers followed
- Incompatibility of the audit software language syntax with the semantics of the data structure models used by the DBMS
- Portability of the audit software package
- DBMS environment is simulated so that integrity features such as concurrency control cannot be checked

USE HOST LANGUAGE EXTENSIONS

Advantages

- Familiarity of the software to the auditor
- Independence of the auditor is partially maintained
- Interface is efficient

Disadvantages

- Self-contained versus host language dichotomy
- Lack of standardization of DBMS
- Incompatability of audit software language syntax with the semantics of the data structure models used by the DBMS
- Portability of the audit software package
- Independence is compromised by using the host language extensions or operating system access routines
- Integrity functions of the Database Manager may be by-passed
- Database definition used by the DBMS may be inadequate for audit software purposes
- DBMS environment is simulated so that integrity features such as concurrency control cannot be checked.

Viewpoint: While much existing audit software does not interface with database systems, a few packages do. These interfaces provide a competitive advantage in the market place which they serve. Auditors are best served by being informed about what packages can work with database systems. Vendors may also be encouraged to improve their products to provide necessary access functions in an easy-to-use manner.

Viewpoint: The advent of a data base management system provides the death knell for most audit software packages. Unless the vendor is committed to expend sufficient capital to provide interface, the onus is on the auditor.

The Auditing Department must have sufficient expertise to provide a "front end" to existing software. This ability is expensive. Programmers with sufficient expertise to program for modern data base packages are in short supply and in great demand. Nevertheless, it is the responsibility of Auditing to provide independent interface to the data base.

In some cases, programs can be written by the systems and programming departments provided that sufficient review by an independent third party is performed and that the programs remain under audit control.

Question 2: In the absence of such an interface, how can the auditor gain access to and manipulate information on a data base?

Viewpoint: Present options appear limited to obtaining a sequential tape or disk file from the client and then processing it using existing audit software. Alternatively, special programs may be written to extract and possibly manipulate data. The former is a reasonably attractive option that is frequently used in practice, but does have the disadvantage of making the auditor somewhat more dependent upon data processing personnel than is true in a non-DBMS environment. The latter solution is generally not feasible because of the time and level of expertise required in order to design and implement specialized programs. In fact, this alternative is probably not economically feasible in audit engagements of less than about 3,000 hours.

No auditor whose employer or client uses IBM 360 or 370 computers needs to do without a database interface. In other environments or with very specialized or complicated structures, the auditor may be able to behave like a regular user for routine information requests. These must be considered as not independent for audit purposes; nevertheless, it may be a very useful procedure.

Viewpoint: Include required audit functions in the generalized language facilities of the DBMS.

<u>Advantages</u>

- o Auditor is not responsible for the interface
- o Compatibility between the language syntax and semantics of the data structures used by the DBMS (except possibly for network based data structure models)
- o Interface is efficient
- o DBMS environment is not simulated

<u>Disadvantages</u>

- o Auditor independence is compromised

- o Self-contained versus host language dichotomy
- o For host language systems, often
  - (a) deficient data definition
  - (b) deficient generalized language interfaces

Viewpoint: Though an auditor wishes independence, he can't have it. He becomes another user in the eyes of Systems and Programming and must depend on programs written to his specifications. In short, redevelop audit software.

Question 3: In working through or with a DBMS, how can the auditor be sure he has been given access to all the records he wants to examine?

Viewpoint: To be sure that all records desired by the auditor have been given to him, he must control the retrieval. (Furthermore, he should create a total file of all records requested to be compared to external data used elsewhere in the business organization.)

Viewpoint: Short of maintaining his own version of the DBMS maintained by resident software experts, he cannot be 100% sure. The auditor is capable of proving record counts, hash totals, balances, etc. to figures maintained by the operating department responsible for the data base.

He cannot, however, be 100% sure that the DBMS has not been compromised and is giving incomplete data back to the user.

Viewpoint: Ensure the conformity of the database to a single database definition.

- (a) administrative aspects
- (b) technical aspects

Develop software which can identify floating or broken chains of data, and data without an existent database definition.

CONFORMITY TO DATABASE DEFINITION

- o Administrative Aspects
  - (a) documentation
  - (b) auditor/DBA interface
- o Technical Aspects
  - (a) only one database definition should exist
  - (b) database definition should be complete
  - (c) DBMS should validate data against the database definition

FLOATING OR BROKEN CHAINS

- Algorithms are needed which can check the pointer fields within a data record to identify floating or broken chains
- Algorithms are needed which can identify data for which a corresponding data definition does not exist

Viewpoint: Although a potentially troublesome area, it is not too much different from the environment which exists today in a non-DBMS environment. In these non-DBMS circumstances, auditors extract records which can be traced to or reconciled with independent sources such as the general ledger. The DBMS environment does add a dimension not present in conventional EDP systems in that a data base administrator has extensive knowledge both of the system and how data is stored and used, and may, unless well controlled, be in a position to perpetrate a fraud which would be extremely difficult for an auditor to detect. On the other hand, a well controlled data base administrator appears to offer control features which are not possible in more traditional environments.

For the present, auditors have no alternative other than to continue extracting data which can be traced to or reconciled with independent accounting records, coupled with insisting upon good internal control procedures over the data administration function. This response clearly rules out accepting data which cannot be traced to independent sources and hence would prohibit use of a DBMS to obtain a listing of say all accounts in excess of 90 days.

Question 4: What controls or features should the auditor look for in evaluating the integrity of a DBMS?

Viewpoint: A DBMS must perform extensive editing of data entering the data base so as to preclude erroneous information from updating a data bank which will be shared by many users. Secondly, a DBMS should provide an effective means to correct and re-enter errors (such as an invalid customer number) which are rejected by the system. It is important to control the occurrence of errors and their subsequent correction.

Viewpoint: The auditor needs some means of ensuring that:

- The manufacturer of the DBMS has adequately tested the DBMS, and subsequent modifications before releasing the DBMS for production running.
- Unauthorized modifications of the DBMS have not occurred within the user installation.

MANUFACTURER TESTING

- Relevance of statistical theory to software testing
- Software development practices of the manufacturer

- o History of use of the DBMS
- o Software certification?

UNAUTHORIZED USER MODIFICATION

- o Availability of system documentation
- o Ease with which the DBMS program code can be understood
- o Availability of systems software expertise within the installation
- o Management and control practices over system software within the user installation
- o Audit testing of critical functions within the DBMS
- o Algorithms for detecting modified code (e.g., comparison of the user package against a manufacturer blueprint, or some kind of hash total checking

Viewpoint: Desirable features of a DBMS which impact its integrity are:

(a) Integration with a data dictionary
(b) A "dump" utility for back-up
(c) Checkpointing for timely restart
(d) Database recovery by optional "rollforward" or "rollback" depending on the cause of the problem
(e) A utility program to restore any part of the data base from the back-up copy
(f) An easy-to-use, flexible, efficient retrieval tool which may be used for diagnostic purposes and/or ad hoc reporting
(g) Minimal application programmer intervention in the management of data base structure information

Similar controls as above, namely--record counts, hash totals, selective field balancing. A log tape showing before and after images is a necessity for on-line, real time processing. The log tape also becomes a factor in recovery/restart, etc.

The ability to define logical data bases (a la IMS) is a super tool. The auditor need only review the logical DBD to determine what action a program can take against a file. This technique restricts access and up-date capabilities.

Question 5: What aspects of checkpoint/restart, recovery, and backup should be of concern to the auditor:

Viewpoint: Auditors should satisfy themselves that adequate recovery features exist in the event of system failure. Additionally, many DBMS systems offer options such as dual logging capability to better insure recovery in the event of system failure. Although batch oriented systems typically did not require auditors to investigate restart and recovery procedures, such is not the case in a DBMS environment. Auditors should be satisfied that, when system outage occurs (as it periodically will), adequate and effective procedures are in place to assure accurate and reasonably prompt recovery.

Viewpoint: The major aspects of checkpoint/restart, recovery and backup that are of concern to the auditor are:

- o concurrency control features
- o what facilities are provided

Viewpoint: Checkpoint/restart, backup and recovery should be tested before an emergency makes it necessary. The relative costs of various recovery techniques should be compared to the losses which might be incurred with increasingly less responsive techniques.

Viewpoint: The auditor must feel confident that the procedures for backup and recovery are adequate. To accomplish this, he is required to test the procedure as he would any major production system. A comprehensive procedure manual must exist showing what is to be done at what time. The auditor must insure that all data is processed and that any hardware malfunction does not impact the ability to process all data.

Along those lines, the auditor is concerned with duplication of master files, the ability to rerun from yesterday's files, offsite storage of master files, procedures for backing up program files, etc.

Question 6: What impact will a DBMS have on the audit or management trail?

Viewpoint: Audit trail is no less important in a DBMS environment than in other data processing environments. The ability to trace transactions from their summary through to detail and vice versa is one which a well designed DBMS should preserve. In those instances where an inability for this to be accomplished exists, the apparent difficulty lies in poor system design rather than in any inherent change in converting to a DBMS environment. Other instances of conversion to DBMS systems indicate adequate planning with auditors has always permitted audit trails to exist in a form which can conveniently be used during the conduct of the audit examination.

Viewpoint: A DBMS has little effect on the audit trail. In a shared data environment, greater emphasis needs to be placed on ensuring that:

- o available audit trail exists
- o a methodology for threat monitoring exists

Viewpoint: The audit trail is being cluttered with less and less paper.

Viewpoint: Dependent upon the design of the system, a DBMS is likely to have a positive result on an audit trail. Correct use of a data base system will provide a trail for the auditor which is likely to be superior to his current systems trail. The accent must be on correct usage as incorrect or incompetent use will befuddle, cloud and obscure an audit trail.

The auditor must have sufficient expertise to determine the adequacy of the trail he will receive.

Question 7: What security features should the auditor look for in evaluating a DBMS?

Viewpoint: Auditors must carefully review controls over the data base administration function, as well as the process by which sensitive data is, first, determined and access to it is, then, restricted. Also important are the procedures by which security violations are detected and promptly investigated by a security officer. Because many data base systems are designed to provide an interactive aid in managing affairs of the company, auditors should evaluate the extent to which data access is restricted to only authorized individuals, reasonable control is placed over the data base administration function, and individuals are denied access to data which would create an incompatible function.

Viewpoint: The DBMS must provide underlying integrity functions to ensure the existence, quality, and privacy of data (Everest; 1974*).

EXISTENCE

- o Backup
  - (a) dual recording
  - (b) dumping
  - (c) logging
  - (d) residual dumping
- o Rollback/Recovery

QUALITY

- o Validation
  - (a) stored data
  - (b) input data
- o Concurrency control
- o Update authorization

---

* Gordon C. Everest, "Concurrent Update Control and Database Integrity," in J.W. Kimble and K.L. Koffeman, eds., Data Base Management (Amsterdam: North-Holland Publishing Co., 1974), pp. 241-270.

PRIVACY

- o Access Regulation
- o Encryption
  - (a) transmission
  - (b) stored data
- o Threat Monitoring

Viewpoint: Security features are the result of considerations of the application being served by the DBMS. The DBMS should be capable of requiring the necessary authorization for anyone to add to, change or retrieve from the database. The DBMS should provide a virtual certainty that all accesses to the database are recorded.

Viewpoint: The auditor should look for:

- o The ability to restrict access by a program to a file
- o Terminal security features such as:
  - logon
  - logoff
  - restart, etc.
- o The ability to define terminals by function
- o Restart and recovery
- o Logging ability
- o Control consoles ability to inhibit a terminal after attempts to logon or process incorrectly
- o Logging function with before and after images

Question 8: What should the auditor's role be in evaluating the impact of privacy considerations or legislation on the design of DBMS?

Viewpoint: The auditor should be knowledgeable about what reasonable privacy considerations are possible. He should understand what privacy features management policy is directing to be implemented. Most important, he MUST know what features are actually being used and whether they are effective. Legislation is simply the public overreaction to situations which private parties have created by failure to act in a prudent manner. It will continue to complicate the already confusing subject of privacy. Effective audit performance requires close contact and frequent reporting to top level management.

Viewpoint: The external auditor will probably be the party responsible to management for ensuring privacy legislation is enforced within the systems of an organization.

The external auditor is responsible to parties external to the organization. The loss of assets which could result from a legal suit over privacy may cause external parties to look to the external auditor for attestation as to the enforcement of privacy legislation within the organization. Interested parties such as the government, socially conscious groups, stockholders, etc., and the organizations themselves may look to the external auditor as an independent party who can attest to the enforcement of privacy legislation, because the integrity of data has continuously been the essence of auditing.

The following three major aspects of privacy legislation are relevant if the auditor is an involved party.

<u>CONTROLS ON OPERATING PROCEDURES</u>

An organization must:

- Take precautions against natural hazards and other threats to the system and its data
- Publish descriptions of its system in a medium which is most likely to be seen by those people who are the subjects of the system
- Establish procedures for responding to inquiries from individuals about their records and for settling complaints about their accuracy
- Keep a log of all users of each person's records and the intent of that use
- Make a person responsible for the enforcement of privacy legislation (the data base administrator?)
- Ensure that data is timely and accurate
- Inform a person if he is a subject in a system

<u>ACCESS RIGHTS OF DATA SUBJECTS</u>

A subject may:

- Examine his own record
- Request correction of erroneous information
- Append a statement to the record if the error is not corrected to his satisfaction

<u>USAGE CONTROLS</u>

An organization must:

- Inform a subject of the intended use of the data, and inform the subject if a new use becomes apparent (implications of this in a shared data environment)
- Use data only for its stated purpose
- Transfer data to a new system only with the permission of the subject, and only after ensuring that the privacy of data will be adequately maintained in the new system

Viewpoint: No involvement. That is a legal not audit problem. It is up to Systems and Programming to insure legal involvement. Audit should only bring both parties together.

Viewpoint: Independent auditors should not be used to assist in monitoring compliance with the myriad of federal, state and local law which govern our society. Although it is reasonable for independent auditors to participate in some compliance, that participation should be limited to situations which have a direct bearing on financial position or results of operations. Because it is not possible for auditors to be conversant in all areas of prevailing legislation concerning privacy, this area should not have the involvement of independent auditors.

6. IMPACT OF GOVERNMENT REGULATIONS


REGULATIONS

6. IMPACT OF GOVERNMENT REGULATIONS

Working Panel Report on Government Regulation

Chairman: Charles D. Trigg

Biographical Sketch

Mr. Charles D. Trigg is Associate Director, National Association for State Information Systems. Mr. Trigg has served as State Comptroller and Budget Director for the State of Missouri. At IBM, he held national responsibility for systems in the finance, tax, and legislative areas of state and local government. He is a member of the National Association of State Auditors, Comptrollers and Treasurers and the Municipal Finance Officers Association. He has testified frequently in Congressional hearings on the impact of legislation in data base areas.

Participants*

James Burrows	Daniel B. Magraw
Charles Burr	Susan K. Reed, Recorder
Robert Caravella	Terrance F. Swanson
Robert Goldstein	

6.1 Scope and Concerns.

The Government Regulation Working Panel interpreted its assignment as follows:

- o To predict which statutes or governmental rules or regulations which now exist or will come into being during the next five years will relate to information systems;
- o To identify which of those will impact data base management methods, procedures, and systems;
- o To make a general assessment of the extent of those impacts with respect to management, technology and cost; and
- o To convert these conclusions to a set of guidelines helpful to top management in making DBMS decisions and, conversely, caution law makers and policy makers on the issues of various proposed policies and regulation.

* Complete addresses and affiliations are in Appendix C

The panel expects a substantial amount of regulation relating to information systems to appear in the next five years originating from statutes and ordinances, rules and regulations issued pursuant to statutes, executive orders, and administrative procedures. It seems clear for at least the next five years that Federal enactments will be the dominant factors in all except those few states with more stringent requirements. The panel concluded that nearly all regulations impacting information systems that are likely to be seen in the next five years are already evident in existing laws relating to privacy and freedom of information. It is anticipated that the areas of impact, nearly all of which are already seen at the Federal level and in some non-Federal governments, will become commonplace in all governments, and toward the end of the five year period, throughout the private sector as well.

Twenty areas were identified in which it is believed that governmental regulations will come to exist nation-wide, affecting both public and private sectors. Ten factors which are part of a total information system, and on which the panel felt the impact of regulations would fall, were selected. The difficulty of analyzing the effect of regulations on systems was increased by the necessity of asking the following questions for each area of regulation and attempting to consolidate the discussion in terms meaningful to data base system managers and users:

(1) Will the regulation impact information systems?

(2) If it does, how does it affect DBMS?

(3) Is the impact on a DBMS generally different from the effect on a non-DBMS information system?

(4) Does DBMS have any inherent advantages or disadvantages in responding to the requirements?

Accordingly, a matrix (see figure 1) was constructed to serve as a basis for analysis. The rows comprise the twenty expected areas of regulation and the columns represent the factors on which management would focus in assessing the impact of regulations. A matrix entry is an affirmative answer to question 1 and an indicator of which factors are affected in answer to question 2. In the COSTS columns the use of two different matrix entries also enables question 3 to be answered. Answers to questions 2, 3 and 4 are expanded more fully below.

After further definition of the areas of regulation, the salient points in the panel's discussion of the impact of regulations on management, technology and costs will be presented.

AREA OF REGULATION	MANAGEMENT				TECHNOLOGY		COSTS			
	A. ORGANIZATION STRUCTURE	B. USER MANAGEMENT	C. EDP MANAGEMENT OPERATIONS	D. ORGANIZATIONAL DISCIPLINE	E. HARDWARE	F. SOFTWARE	G. HARDWARE	H. SOFTWARE	I. PEOPLE	J. MISSED OPPORTUNITIES
1. SYSTEM CERTIFICATION	X	X	X	X	X	X	Y	X	X	
2. STANDARDIZATION OF PROTECTION OBJECTIVES		X					Y	Y	Y	Y
3. SUBJECT ACCESS RIGHTS	X	X	X			X	X	X	Y	
4. NOTIFICATION OF PRIOR RECIPIENTS							Y	X	Y	
5. DATA COLLECTION LIMITATIONS									Y	
6. LIMITS ON INTERRELATING		X				X		Y		Y
7. UNIVERSAL IDENTIFIER									Y	Y
8. DATA RETENTION						X		X	Y	
9. CONSENT FOR DATA USAGE									Y	
10. ACCURACY, COMPLETENESS, TIMELINESS		X				X			X	
11. ACCESS AUTHORIZATION	X	X	X	X	X	X	X	X	Y	X
12. CORPORATIONS AS INDIVIDUALS										
13. NON-PERSONAL DATA					X	X				
14. CONTINUITY OF OPERATIONS		X	X		X		Y			
15. DBMS STANDARDIZATION		X	X			X		X		X
16. AUDIT TRAILS		X	X				Y	X		
17. DEDICATED SYSTEMS	X					X	Y	X	Y	X
18. PROGRAM STATUTES				X					Y	
19. STANDARD DATA ELEMENT DEFINITIONS, CODES								X		
20. FREEDOM OF INFORMATION		X						X		

X - Impact

X - difference in cost under DBMS and non-DBMS

Y - no material difference in cost under DBMS and non-DBMS

Figure 1. Impact of Regulations on Data Base System Factors

6.1.1 Explanation of Matrix Rows as Areas of Regulation. Each explanation begins with the number of the matrix row.

1. System Certification - The operator of a data base system must assure that his system complies with all of the specific regulatory requirements. This might involve the use of an external and internal auditing organization.

2. Standardization of Protection Objectives - All systems will have to provide a common level of information protection. This does not necessarily imply the use of common protection techniques.

3. Subject Access Rights - Individuals will have the right to find out if they are the subjects of data in a system and, if so, what information about themselves is stored. They will also have the right to have errors corrected in their records.

4. Notification of Prior Recipients - When an individual has an error in his record corrected, the system operator will be obligated to notify past recipients of the error. This may be an automatic, blanket notification of all past recipients, or a selective notification at the request of the data subject.

5. Data Collection Limitations - Organizations will only be permitted to collect information from individuals that is relevant to the functions of that organization. In general, the consent of the data subject will be required.

6. Limits on Interrelating Data - There may be restrictions placed on the interrelating of information from different files or systems.

7. Universal Identifier Use - No universal identifier will be established within the next five years in the U.S. It is possible that the use of common identifiers between systems will be explicitly prohibited.

8. Data Retention - Specific maximum retention periods will be specified for certain kinds of unfavorable personal information. Minimum retention periods may be specified for other information such as record usage logs.

9. Consent for data Usage - The informed consent of the data subject must be obtained before information about him may be used, except for uses specifically authorized by law.

10. Accuracy, Completeness, Timeliness - Organizations maintaining personal data must keep that data in a sufficient state of accuracy, completeness, and timeliness that fairness will be ensured in any decision making based on that data.

11. Access Authorization - The system must prohibit all data accesses except those specifically authorized.

12. Corporations as Individuals - Corporations will have the same rights with respect to information about them that are currently granted to natural persons under the 1974 Federal Privacy Act and similar laws.

13. Non-Personal Data - Data not covered by the various privacy laws must also be protected against loss, alteration, or improper disclosure.

14. Continuity of Operations - Organizations must ensure that they are protected against disruption of their normal operations as the result of loss or damage of data.

15. DBMS Standardization - A standard for data base management systems may be established by the official standards organization, or de facto, by decision of a major user such as the Federal government.

16. Audit Trails - It will be necessary to maintain a log of changes and disclosures of data. This is needed as an aid to maintaining data integrity, for use by the system auditors, and to enable data subjects to find out about the usage made of their records.

17. Dedicated Systems - Separate data processing systems may be obligated by certain applications. This will impact the extent to which the benefits of a DBMS can be realized.

18. Program Statutes - The individual laws and regulations governing various organizations may include provisions relating to information processing tasks.

19. Standard Data Element Definitions and Codes - De facto standard data element definitions and codes may be established through their adoption by a major user, such as the Federal government.

20. Freedom of Information Acts - Many governmental bodies are subject to laws authorizing a wide range of citizen requests for information. This places additional demands on their data management facilities and may, in some cases, conflict with protection provided under various privacy statutes.

6.1.2 Explanation of Columns.

6.1.2.1 Analysis of Impact on Management. It is clear that implementation of a data base management system has organizational implications. In order to comply, in an efficient and cost effective way, with laws and regulations currently contemplated, an argument can be made that "CONTROL" or "PRIVACY-SECURITY ENFORCEMENT" should be centralized administratively. It is not obvious where this function should appear in the organizational structure, but such administrative responsibility must necessarily be close to and involved with the systems and programming technical staff while at the same time being high enough in the hierarchy to produce effective enforcement as well as to affect general D.P. policy.

The existence of a DBMS will make the implementation of regulations and laws more uniform throughout the purview of the DP user community (public or private) and substantially simplify the job of enforcement. The panel believes that DBMS will be able to respond to changing and new regulations and laws more flexibly and easily, thereby reducing the need for a technical manpower investment in each new requirement. Thus, DBMS and the thrust of expected legislation and public policy seem to compliment each other in terms of centralizing responsibility for managing the data base AND enforcing the laws and regulations, which will be promulgated in any case. This seems to impact on the debate between disbursed data base advocates and those supporting the philosophy of centralized data processing.

An examination of the impact the predicted regulations will have upon the management structures of both users and data processing groups indicates that their responsibilities will probably increase in proportion to the number of new rules and regulations under which they must operate. It would further appear that any penalties imposed for failure of agencies to comply adequately with the regulatory system would fall most heavily upon the individuals in these groups.

In view of the pressures which will be exerted upon them, it is reasonable to expect that they will welcome any technique or system which could ease their tasks. Data Base Management Systems should enable them to design and control systems more easily, which would conform to regulatory requirements.

In fact, without DBMS techniques control procedures would become exceedingly difficult to establish and cumbersome to follow. Specifically, DBMS will facilitate procedures for certification and standardization of data systems. Its use will also simplify the control of data accessibility and it will ease the task of assuring the accuracy and timeliness of data. DBMS provides both the users and data processing managers with a tool which will expedite their compliance with the anticipated regulations.

The general public view that DBMS can either cause or assist the unwarranted interrelating of data may be significantly reduced by appropriate publicity given to management policy and the stringency of controls over management as well as users. This sort of limitation can be more effectively and efficiently enforced under a DBMS than if the data is scattered among several different non-DB systems (even though the logistics of interrelating data in separate systems are more difficult).

Similarly, access limiting rules and regulations force organizational discipline and require specific administrative control of access (via policy and software). Again, such discipline is enhanced by DB systems. User management will be forced to specify access authorization by individual and by data element, ultimately.

An interesting interplay of factors indicates that the increased flexibility for interrelating data and for browsing in DBMS requires more stringent definition of accessability by data element and more stringent and precise audit and control of the "user" in his utilization of the system. The extended capability of DBMS both requires and enables these functions.

6.1.2.2 Technology. Many of the regulations which are awaited will have impact on the availability, use and safeguards of data in current or planned information systems; in fact, all statutes passed to date apply to data--not to systems. But it is in the systems that procedures must be implemented, and certified as adequate, to meet the legislative aims.

System managers, especially those with on-line access, either local or remote, have an extensive task before them in certifying that the hardware, software, and procedures of implemented systems will indeed carry out the regulatory intent. A manager of a system which is built upon in-house developed structures, such as locally developed mini-DBMS's, special hand tailored higher-level or machine language, code, etc., will be working alone when he comes to test and certify his system. However, if he builds his system upon a DBMS or a standard package which has an extensive user community, he will gain the benefit of a cooperative effort which can lead to certification of the system. Participation in such a group to share discovered defects, emergency procedures for fixing them, and, ultimately, procedures to correct and extend the DBMS will significantly reduce the risk and cost of implementing the intent of the directives.

It is probably not easy for legislators to understand why errors in data can occur in the large systems implied by the use of our current data base technology. While no one intends to create or accept errors, the current state of the art in specification and testing of programs cannot handle the complexities of our current (and even our first generation) systems. Nevertheless, it should be noted that there is considerably less bad data in automated systems than was contained in manual systems.

To deal with some of the specifics of current or impending legislation, a DBMS, if used, will require capabilities which may not have been previously needed, e.g., in the Privacy Act of 1974 there is a requirement for Federal agencies to allow an individual access to information pertaining to himself which is in the system and which is specifically accessible by a common or unique identifier. In addition, when such information is proved to be incorrect, it must be corrected at the request of the individual. To accomplish this economically when there is a significant number of inquiries, with adequate controls to differentiate such transactions from those of the agency which is using the data in the conduct of its mission activities, the DBMS must have an efficient batch inquiry and update capability.

Although one of the requirements usually imposed by privacy legislation is that the data in the system be accurate, complete, and timely (current), there is very little except edit checking for reasonableness that can be done once the data is in the system. However, a DBMS does offer an order of magnitude increase in maintaining the integrity of data once captured. The problems associated with integrity, including recovery and restart, are significant and their resolution is not trivial either to design or to implement correctly. Most creators of DBMS have attacked these problems. Their current products represent some of the best ideas in design and have been tested by usage. Thus, there is an advantage in using a DBMS to preserve the accuracy and availability of data. This integrity feature also increases an agency's ability to have its data processing available for mission support at any time, i.e., accidental mishaps will cause fewer lengthy outages of service.

In the Privacy Act there is a requirement to guarantee that the data is only used for authorized and announced purposes by personnel who have individual (or sub group) authorities to access the data. For administrative control of such use and access, it is essential that a central authority have a viable and credible capability for enforcement. A DBMS, because it must provide an essential mechanism for controlling access to data, contains an ideal place for capturing, inspecting, and authenticating all requests for access, either by individuals or for specific uses. While this feature may not be available in all DBMS, it should be locally implementable; and if the DBMS has a wide user group, such features can be well checked out long before a home-grown access control subsystem could be.

Although the technologist-managers have been striving for some years to define an acceptable, if compromise, standard DBMS, along the lines of the standardization of COBOL, it is considered premature for the Federal Government at this time, or in the near future (3-5 years), to endorse a procurement policy requiring that mainframes bought by the Government have a DBMS which meets precise specifications. Such a step would have far reaching effects. First, all main frame manufacturers would have to decide to implement/acquire such a system or withdraw from direct selling to the Federal market. Presumably a third party could develop such a software system for a specific set of hardware and bid on Federal specifications. This is, however, not likely due to the bid costs of preparing for live test demonstrations, etc. Second, if industry did decide to prepare to bid on Federal specifications, most manufacturers would probably not continue development of alternative DBMS. A DBMS is a very expensive system to build, maintain and extend. If the Federal specifications were based on extensions of the work of the Data Base Task Group of CODASYL, a few manufacturers would continue on their current course. This would entail a technical evaluation by each manufacturer of its customer investment.

6.1.2.3 Costs. The emphasis in this discussion is on distinguishing between what will happen if a DBMS is selected instead of another system. It is an acknowledged fact that governmental regulation will increase the cost of an information system. The concern here is to determine whether implementation of government regulations will cost more or less under DBMS as compared to a non-DBMS approach.

For this reason an additional matrix entry was employed in the COSTS columns only (see figure 1):

x - indicates that costs will vary between DBMS and non-DBMS

y - indicates that there will be no material difference in cost between DBMS and non-DBMS

In nearly all cases where the entry in the COSTS columns is an X, i.e., when there is a difference in cost, the advantage lies with the DBMS approach. Additional hardware requirements which might be imposed by regulations will generally be less under DBMS for two reasons. First, when individuals exercise their right to access data and correct it, DBMS can access multiple files faster and less expensively. Second, when regulations require that access to data be controlled, which is costly in any system, it can be done in less time and less expensively under DBMS because of the centralization of the control function in a single program module.

With respect to software costs, eight of the ten items marked X in the matrix would be less costly under DBMS because they make use of the inherent capabilities of the DBMS. The item of DBMS Standardization would, of course, only apply to DBMS. The costs of DBMS might be greater under a decentralization requirement; the main bulk of such costs would go for non-DP activities, such as investigations. (N.B. it is emphasized that these conclusions relate only to the impact of Government regulation on systems. Other factors important to the DBMS decision have not been considered here.)

Personnel costs would differ under DBMS vs. non-DBMS only when concerned with insuring data accuracy, completeness, and timeliness; in such a situation DBMS would be only marginally less costly.

Any group considering costs would be remiss if it did not also consider that there is a cost in "missed opportunities," i.e., activities denied an organization because of limitations imposed on data transfer either by statute or by expense. Imposing too costly an inter-system data security standard may result in an inability or unwillingness to participate in programs which require the standard. Prohibiting the use of a universal identifier or the interrelating of data will likewise preclude data transfers. A DBMS standard would enable DBMS to control access better--a benefit rather than a missed opportunity. This apparent advantage notwithstanding, imposition of a DBMS standard could have a negative effect on overall performance in many cases. Most unequivocal of all is the statement that if decentralization becomes a requirement, great opportunities will be lost.

6.2 Conclusions

During the course of the analysis of the impact of Government regulations on data base management systems and the ensuing discussion, a number of general conclusions, some almost axiomatic in nature, were reached by the panel:

1. Existing and proposed regulations will impact organizations whether or not a DBMS is used.

2. State and local governments should have standard privacy/security regulations if they have a requirement to exchange data. In the absence of these standards, the Government runs the risk of not being able to exchange data because its privacy/security requirements are either too stringent or inadequate to permit exchange with the target government. This implies that there will exist some entity or some way for these governments to certify that reasonable precautions exist to safeguard the transfer, use and storage of the data. This does not imply that the same data base management system must be used--only that "consistent" levels of protection must be provided.

3. The decision to implement DBMS may be favorably impacted by existing and proposed regulations. The use of DBMS offers organizations a flexible alternative to respond to changing as well as new regulations. In the absence of DBMS, new requirements may have a costly impact by forcing systems conversion and/or the development of systems enhancements which were not originally addressed as an integral part of the system design. Carrying that idea even further, some regulations may prove to be prohibitively costly to implement without the use of DBMS technology.

4. The possible regulatory requirement which could unnecessarily over-burden an information system is the need to notify all previous recipients of data on a given subject of subsequent changes (additions, deletions, modifications) to the record. The problem may be compounded if the primary custodians of the data have disseminated it to secondary and tertiary users. Notification of previous recipients should be required only if they are specifically named in writing by the data subject. This would alleviate the burden by requiring notification of only those users about whom the data subject is concerned.

5. With respect to organization structure, those organizations whose prevailing management philosophy encourages centralization of control will probably be more amenable to adopting the DBMS approach. Organizations which emphasize decentralization of accountability should approach the DBMS decision with an awareness of the possible broader implications on its approach to management.

6. Within DBMS is a software function, known as the data base manager, which serves as the natural point to control as well as maintains surveillance of access to multiple files, data elements, programs and terminals. Manual systems are not capable of handling system resources nearly as efficiently.

7. The inherent flexibility and responsiveness of the DBMS carries with it some attendant problems including--

(a) the need to impose more stringent administrative controls on the DBMS operating environment.

(b) the risk of data base destruction given the dependence of numerous application systems on the single source of data.

8. If corporations are eventually included as "individuals" within the scope of the Privacy Act of 1974, or if the security of non-personal data is regulated, there will be no additional impact on DBMS that has not been previously discussed.

9. Since the current Federal law requires a roster of information systems and their basic characteristics to be published in the Federal Register, freedom of information should not require the notification of data subjects that information exists in a particular file about them. This approach would be so prohibitively expensive as to destroy the ability of Government to function. Rather, systems should be able to respond responsibly to initiatives of possible data subjects. Obviously, the DBMS approach would facilitate such a policy.

10. Policy makers and law makers are cautioned that prohibiting, or limiting, the use of a universal identifier is primarily of psychological value. It provides the illusion that without universal identifiers data cannot be interrelated. The fact is that their absence only makes more difficult and costly the task of legitimate data correlation as well as increases the cost of complying with freedom of information laws and policies.

In addition, the lack of a universal identifier greatly complicates the problem of insuring the accuracy and completeness of data (thereby increasing cost or increasing error levels). Laws should be established which control the interrelationship of specific types of data. Constructing hurdles to prevent misuse of data by making it too costly to correlate will not solve the problem and is counterproductive to the efficient operation of Government and industry. In the end, the taxpayer/consumer will have to pay this unnecessary cost.

7. DATABASE TECHNOLOGY--PRESENT AND FUTURE


TECHNOLOGY

7. DATA BASE TECHNOLOGY -- PRESENT AND FUTURE

Working Panel Report on Evolving Technology

Chairman: George Dodd

Biographical Sketch

Dr. George Dodd is Assistant Head of the Computer Science Department of General Motors Research Laboratory. Previously he was a senior research engineer with the Laboratories and an instructor at the University of Illinois. He received his Ph.D. from the University of Illinois. His research and professional interests are in computer graphics, data bases and virtual memory time sharing systems. He is the East Central Regional Representative for the Association for Computing Machinery and also served on their Long Range Planning Committee and their Government Reorganization Committee. He served as a member of the CODASYL Data Base Task Group and has written and lectured extensively on the subject of data bases.

Participants*

Charles Bachman	W. Frank King, III
Richard Baum	Eugene Lowenthal
Susan Brewer, Recorder	Jack Minker
Emil Broadwin	Michael Senko
James Fry	Edgar Sibley
Michael Kaplan	Dennis Tsichritzis

7.1 Introduction

The charter of this panel was to examine the evolution of technology as it affects data base management systems (DBMS). In particular, the panel members were instructed to examine the technical areas discussed herein and to prepare recommendations on how the manager of a computer installation should react concerning the development of data base systems over the next five years. In addition, the panel examined the directions of technological evolution over the next ten years and summarized the work to be undertaken to achieve reasonable progress.

This panel included members from the user community, academia, CODASYL, manufacturers of computer equipment and industrial firms. This spectrum provided a broad view of the overall directions we expect data base management systems to take.

---
* Complete addresses and affiliations are in Appendix C.

Four categories of topics were discussed. They are:

1. USABILITY which includes data base specification, use of programming aids, data base tuning, availability of data bases, error recovery and data independence;

2. DATA BASE ARCHITECTURE and distributed data base systems;

3. NEW FUNCTIONS which include data base models, relational inferences, natural languages and data base semantics; and

4. MISCELLANEOUS which covers standardization and research financing.

7.2 Major Conclusions

7.2.1 Data Base Usability. Very few facilities exist for developing a statement of a data base design and a statement of how and when restructuring would expand DBMS use or improve performance. However, specific areas such as the ISDOS project at the University of Michigan will show some progress. With the current work in structured programming and design, an upsurge of effort in the design area will occur, although useful products are unlikely for the next three to four years. The panel noted a number of selective tools that collect statistics and simulate performance for various data base systems. Though not universal tools by any means, available technology can provide them. Users of data base systems should require that their vendors provide better statistical measuring tools, simulators, and benchmarking facilities so that they can determine the performance of a data base system before they implement a particular application.

In the area of data base tuning, the panel sees, within the next five years, increased capability to tune manually the data base without having to rebuild it. Tuning will consist of a collection of manually initiated operations executed entirely by the data base management system. These operations will establish new data access paths, add fields of records, install a new access method, etc.

In this same time frame, facilities for recovering data bases from failure and increasing their availability will improve vastly. The panel notes that need will cause the development of rules of thumb to aid the data base managers in evaluating tradeoffs for various levels of recovery. Automatic recovery aids that keep an application data file consistent will also be available in this time frame. User community pressure will cause these improvements which will use existing technology.

The term "data independence" implies that application programs are independent of certain changes to the data used and that new program functions can be added without affecting existing applications. Languages and programs will become more data independent during the next decade. While current systems lack physical data independence, the next five years will bring considerable progress. Cost benefit tradeoffs might indicate that certain types of independence are very expensive and the user should assess these based on their merits and on his needs.

7.2.2 Data Base Architecture. The panel expects to see new types of data base architecture. These types will include: front end processors more closely related to the storage hierarchy and special stand-alone computing systems to do processing of relationships and to permit on-line access to very large data bases. This type of hardware development will parallel the type of evolution seen in the communications area, where most of the communications functions have been removed from the central computer and placed in a peripheral communications computer.

The physical division of a logically integrated data base over several distinct computing facilities is called a distributed data base. Though a technology only in its infancy, the panel expects to see commercially available distributed data base systems in vendors' product lines within five years. The systems are already becoming cost effective in certain specific applications.

7.2.3 New Functions. In the area of data models and supporting languages, the panel notes an era of inventiveness. A number of languages and models either exist or are being proposed. Each of these models has proponents who point to advantages for their particular 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.

Data base systems will become much more intelligent. That is, the user will describe a problem and the DBMS will use the problem statement and information in the data base to infer the solution though that information was not stored specifically in the data base. The same techniques will allow us to have a more natural language approach to data base queries.

7.2.4 Miscellaneous. The panel expressed a concern about the effects of standards on evolving technology. Each standardization effort should be examined on its own and a solution determined on the merits of each proposal.

A tremendous amount of new research is required to achieve the goals set forth in this study. Joint industry-research study projects should be initiated to stimulate this work.

7.3 Data Base Usability

7.3.1 Introduction. Current data base management systems design is an ad hoc process. Consider the methodology:

1. Survey the users of the proposed system to determine the significant transactions to be processed, the types of reports to be generated, and their data base needs.

2. Utilize the survey to propose a potentially satisfactory logical and physical data base structure. (A _logical data structure_ presents the user's view of the organization of his data. It most nearly reflects his problem statement and the way the items of data would be used to solve his problem. The _physical_ or _storage structure_ is the internal organization of the data in the computer memory and on storage devices. The physical structure generally differs from the logical organization to improve operational efficiency. The description of the overall logical and physical storage structure is called a "schema" or "plan." Each user of the data base may have his own view of the data base dictated by his data base update, performance and security needs. The view each user has is called a "subschema.")

3. Implement the system and load it with data.

4. Use the system while gathering statistics about it.

5. Use these statistics to design and implement improvements in the procedures or the data base. This often results in the restructuring of the data base.

Serious deficiencies plague all of these steps, except number 3, as done today. We will now examine deficiencies in three areas.

7.3.1.1 Specification of Data Base Requirements. The data base design process lacks the ability to formally specify the problem requirements to be handled by a data base system and by the computer system encompassing it. In particular, the entire process is one of trial and error. The meaning of the data base is embedded in the programs and the reasons and effects of restructuring the data base are lost in a series of program modifications. Also, as a data base system becomes more integrated, added interrelations often confuse the original intent and structure.

To help overcome these difficulties a number of groups have begun developing formal languages and graphical approaches for specifying the problems to be solved by a data base. At least one of these involves a structured approach in which the system is supplied a stylized description of the data base. The data description includes types of data to be used; storage, retrieval, and update patterns; and type of output reports. This description provides a data base structure definition which can then be used in a COBOL, PL/I, etc. program. In the long range this approach should produce a good, and perhaps optimized, data base description and the entire set of problem solving programs. Later, a changed problem statement could either attempt to use the existing structure or alter the data structure to achieve a new optimal structure.

7.3.1.2 Data Base Tuning. Tuning a data base includes two concepts: improving the performance aspects of the physical storage structure, and applying the usage statistics to seek improvement by changing the quantitative or qualitative aspects of a logical data structure.

Available tuning tools permit the data base administrator (DBA) to affect data base performance via:

1. Modification of the logical schema in those cases when, for example, the data definition itself can contain language statements to build new access paths in order to optimize performance. Modification of the logical schema becomes very expensive if it invalidates existing application programs.

2. Modification of the physical schema beneath an unchanged logical schema, such as using rings instead of pointer arrays to represent a set. Modification of such a physical schema has little impact on application programs.

3. Reorganization of the underlying structures; such as compressing free space, bringing related records together, or rearranging records to minimize deadlocks. Again this should not affect user's programs.

Despite such opportunities, data base tuning has weak support because the DBA lacks reliable guidelines for using the available tools which are often restricted to special systems. Mechanisms used today to improve the operation of data base management systems include:

1. Tailoring - Tailoring is ability to reconfigure the DBMS program itself. Such reconfiguration can involve rearranging program overlays to group together logically connected programs and to move low use or optional data base features into separate program groups.

2. Preprocessors - A program which translates a high-level language into another high-level language in order to provide processing for specialized language features. Preprocessors can provide privacy locks requiring satisfaction prior to compilation or gather statistics on the use of certain language features.

3. Utilities - A program executed independently of the data base management system to convert a data base from one data form to another, to condition input for insured consistency, to sort a file, etc.

4. Statistical Measurements - Other forms of utilization guidelines include facilities to estimate a data base system's space and time requirements so that a data base administrator can estimate the type and capacity of hardware required to operate the system and the organization of software required to use the hardware efficiently. Performance statistics generated by the DBMS need a presentation form usable to the DBA in order to assist him in decisions concerning reorganization and subschema modification.

7.3.1.3 Data Base Availability and Recovery. One of the most pressing problems facing the DBA is to assure that the computer operating system and its data base will be available for problem solving. Three levels within the computer system impact the availability and recovery of data.

On the first level, data entering the computer must have assured validity and fit within guidelines for permitted data values in the data base (see Data Semantics section). Also, the data must have "quality" so that missing data can be handled and erroneous data can be tolerated and accounted for. The nature of a data base's data may require redundancy. In situations of geographically distributed files, a high degree of redundancy may result. A data base structured in one hierarchy of a computer system may not need as great a redundancy. If the operating system should fail, redundancy in the data base often permits recovering the data without excessive effort.

The second level important to recovery and availability is in the operation of the data base programs. Here we provide capabilities for checkpointing the data base system by periodically copying the data base so that in case of failure we can restart and continue. Some systems record each transaction against the data base so that the system can recover up to the last entered transaction. Other systems make a "back-up" copy of the data base periodically and, when an error occurs, restart at that back-up point. The degree and amount of recoverability depends upon the type of problem encountered. User termination of a transaction in mid-stream, a disk crash, or a memory crash -- each presents needs for different types of recovery.

At the third level of importance rests the need to recover from failures in the hardware or operating system. The processes needed for handling concurrent access of the data also fall here since the operating system almost always handles these. Index and data flow coordination must occur in the DBMS programs so that if an element fails, enough indices and information remain to resume operation with minimal effect.

7.3.1.4 Five Year Operational Outlook. The state of the art for data base design reminds one of that for building bridges a century ago. At that time engineers estimated the load that the bridge would bear, tried to make sure that enough steel and structural support would be provided to hold the load--and hoped. Not until the twentieth century when "strength of materials" became a science could we design bridges to withstand earthquakes, wind and water. We build our data bases now much like we built our bridges one hundred years ago. We will very slowly evolve to a better design methodology.

1. System Development Aids. We expect that the filling in of gaps in system development aids will begin in five years. The most harmful gaps are in documentation areas. Vendors will experience an increasing demand for documentation. More preprocessors and utilities will be built.

2. Data Base Tuning. We can reasonably hope to see provided in varying degrees over the next few years, an increasing capability to tune "manually" the data base without literally rebuilding it, i.e., a broader range of manually initiated operations which are executed entirely by the DBMS. This will be accomplished by:

(a) Moving performance related constructs out of the logical schema and into the physical schema. We must minimize the motivation of the DBA to "contrive" the logical schema only for the sake of performance.

(b) Making the physical schema transparent to the user.

Both (a) and (b) can be achieved entirely without stretching the state of the art. Progress in the area of physical data independence (see the section entitled "Data Independence") is extremely important - the tuning tools discussed here will not be exploited unless user programs and user habits can be insulated from the tool's effects. Improved automatic statistics gathering by DBMS will emerge, and vendors will supply evaluation and analysis programs to assist the DBA in interpreting these statistics. Although tuning itself will continue to be manually initiated, the DBA will have better information to work with.

3. Data Base Recovery. At the present time, all data base systems have some limited facility for data base recovery. Some of the most elegant techniques are found in the systems provided by MULTICS. Data base managers have an immediate need for rules of thumb to evaluate trade-offs for various levels of recovery. We expect these to be available within five years. Currently available automatic recovery within the context of an application permit some systems to "back out" of a transaction if an error occurs. Future recovery aids will enable the data base manager to assess the damage and inconsistency of the data base.

4. Summary. The improvements discussed above will come about because of pressure from the user community and will use existing technology. This panel expressed concern, however, that existing tools were not more widely used. Wider use would substantially improve existing data base systems. We recommend that data base systems users demand more from their vendors: more tools to measure data base performance and more tools to help provide backup and recovery. These tools can be made available within today's technology.

7.3.1.5 Ten Year Research Needs

1. Usability: An Epistemic Assessment. Although data base useability constitutes a vitally important area for the future of data base design, it presents problems so complex that they are even hard to state. Very little data exists on the definition and use of current data base systems. We don't know how data bases are being defined. We don't know how they are being used. We don't know the average depth of trees or length of chains or queries per minute or updates per month. We don't know the types of reports being prepared and we don't know the growth rate of data base systems.

To develop an operational requirements language requires knowledge of the items described above. This data will be a long time in coming because much of it is peculiar to specific data base systems. However, we do expect this type of information to be available within the next five years. The long term research need is to collect this data and reduce it into a form so that language, data base, and system designers can use it to develop the requirements language of the future.

Similarly, the development of better recovery aids needs knowledge about how we do it today. Most available recovery aids perform a blanket recovery. We need a finely tuned recovery mechanism to handle special case situations without affecting continued operations of other operating system users.

The panel anticipates intensive investigation into the gradual replacement of manual techniques. The cornerstone of this trend again is the "user profile" which is (at least) a detailed model of the types and frequencies of actual or expected transactions against the data base. Once we have linked the DBMS itself to the user profile, new automatic tuning features will appear in approximately the following order:

(a) Automatic update of the user profile in response to actual job load, i.e., as a byproduct of statistics gathering.

(b) Automatic monitoring of performance - the system itself will detect when a data base reorganization or internal schema modification will improve performance and alert the DBA.

(c) Automatic analysis of the appropriate remedy - the DBMS will not only determine the action required, but also suggest specific remedial steps.

(d) Dynamic Tuning - armed with all this perception and analytic power, the DBMS will actually modify the storage structure in "background" mode in those cases amenable to a remedy carried out piecemeal during periods of relative quiet, e.g., (1) the DBMS will physically delete, when available time permits, a record that was earlier only marked "deleted" and (2) for a set occurrence which appears to experience a great number of owner accesses from members, the DBMS adds a link to owner.

7.3.2 Data Independence. A much-used term in the data base field, "data independence," roughly implies that application programs are unaffected by certain changes to the data they use or by the use of the data by new application programs. In this section we will more fully discuss this term, indicate the importance of data independence to decision makers in the EDP field, and give limited opinions about the future of data independence in data base systems.

Users of the concept of data independence often intend both the physical and logical aspects of data base systems. The panel first separated data independence into these two areas.

Physical data independence means that application programs remain unaffected (except for performance) by changes made to the physical storage structure. Examples of this low level data independence are the ability to change (1) the placement of disk packs on devices, (2) the placement of data on disk packs, (3) the blocking factor of the data sets, (4) the method used to access the data set, (5) the set of indexes used to access the data, and (6) the types and implementations of pointer chains used to represent associations among data items. Current technology already provides a rich range of capabilities to store data with considerable physical data independence.

The other type, logical data independence, has two important aspects: (1) the ability of a DBMS to support different viewpoints of the same data base schema (subschemas), and (2) the ability of the DBMS to allow modifications to the schema without impacting existing applications.

7.3.2.1 Present Technology. Most current technology systems provide the DBA with the ability to define certain schemas of the data base which are not direct maps of the stored representation. For example, in IMS the DBA can define logical data bases in terms of the physical stored data bases by either pruning physical data bases or by interrelating several physical data bases. In DBTG-like systems the subschemas are a subset of the schema and serve a similar interrelation function. However, a subschema may make selected associations between records or segments either visible or invisible to different applications programs. In addition to controlling the associations in a schema, systems allow the DBA to subset the records (by field and even by field value) which may appear in a subschema.

The ability of a DBMS to support schemas substantially different from the stored representation provides a real measure of the value of the system, for this ability can strongly affect the maintainability of application programs and the ability to tune the system as performance requirements become known or change. A simple example may help to clarify this important issue.

Consider an application which must update the master inventory data base depending on a daily transaction data base; i.e., change the inventory quantity at the end of the day to reflect shipments and receipts. A natural approach would assume that both the master and transaction files are sorted by item number. Imagine, however, in our example that the schema view of the transaction file reveals a chronological order. Then the application program must have logic to search repeatedly the entire transaction file looking for all applicable transactions to update a specific item. The program logic becomes complex because of the file ordering and this complexity affects maintainability. If, later, efficient performance became important, the obvious tuning (maintaining the transaction file physically sorted by item number) would not increase the overall performance without a change in the program logic.

The requirement to support a variety of schemas is independent of the particular data model used. For each data model the DB Administrator should know whether the system will support multiple schemas; i.e., in a network model system what class of subschemas can be defined?

The other important aspect of logical data independence is the ability of the DBMS to allow schema modification without impacting existing application programs. Current applications should continue to run even though we changed the length or type of a data item, added new fields to records, added new associations (sets) between records, or added new record types altogether.

Except in a completely interpretative system, these kinds of changes may require recompilation of the program or respecification of the mapping of the stored data representation.

Again, we should stress the importance of a DBMS's ability to accommodate these kinds of changes, especially in terms of program maintenance costs.

7.3.2.2 Five Year Operational Outlook. The Evolving Technology group expects the various aspects of data independence described will develop in an evolutionary manner during the next decade. While current systems do not have physical data independence which is completely separate from implications on the supportable logical views, the group feels that this area will progress considerably in the next five years. DBMS packages, now available, exhibit a very high degree of physical data independence and these will continue to be developing in the five year period. In general, however, cost benefit tradeoffs might make certain types of independence very expensive.

The more difficult support of logical data independence will develop gradually in degrees over a 10 year period.

7.4 Data Base Architecture and Distributed Data Bases

7.4.1 Introduction. The problem of architecture for a data base system resembles that of a building. Given the bricks, glass, carpeting and utility services, the architect designs a building to be functional, economical and meet the users needs. Similarly, we are faced with the problems of organizing the hardware, software and storage of the computer system so that it can economically save, retrieve and manipulate the data base to satisfy user needs. Many variables influence data base system architecture: the size of the data base, amount of available storage, the degree of interconnection or integration in the data base, the speed at which functions are to be performed on the data base, the geographic distribution of the data and the relative frequency of the functions being performed. The most typical objective of considering new or modified architectures is the improvement of cost/performance ratios, although less tangible enhancements, such as data base privacy/integrity, may motivate the architectural design as well.

Influencing the architecture is the degree of data independence (see data independence section), the type of model used to represent the relations of the data (see section on data models and languages), the hardware organization and the degree of distribution of the data base. This section will deal with the latter two issues.

7.4.2. Hardware Organization. Architectural improvements will take place in response to increasing demand for lower costs and increased throughput, capacity and reliability. Hardware improvements will permit existing DBMS to perform better with relatively minor changes, e.g., accepting a higher speed disk. The end user and the DBA will be insulated from these changes.

Examples of these improvements include:

A. Advances in storage technology, e.g.:
   - o large, low cost/bit random access or block oriented memories (e.g., photodigital or optical stores)
   - o memories with bit costs similar to disk but exhibiting much faster access times (e.g., bubble or electron beam)
   - o extremely dense, fairly low cost disk units

B. Transparent storage hierarchy managers, e.g., IBM 3850, CDC 38500.

C. Transparent improvements in processor technology. Higher speeds, greater reliability and lower cost of central processors will have desirable effects on existing DBMS performance, but not as dramatically as that derived from storage improvements.

Adaptive DBMS improvements will also emerge which manipulate stored data in novel ways to exploit fully architectural improvements. The data manipulation language need not be modified to exploit this, so end users are not affected, but the DBA may be confronted with a new set of tradeoffs and tuning tools.

7.4.2.1. Five Year Operational Outlook. New DBMS will probably be developed within the next five years to exploit the use of dedicated data base processors. An example of this is a "backend processor" which is connected to the conventional host or mainframe computer. The data management function is distributed between the mainframe, which handles the user interface, and the backend, which manages the storage interface. The main benefits will be (1) reduced inefficiency, penetrability, and vulnerability of the general purpose hardware, operating system and file management system of the host, (2) unburdening of central memory, CPU and channels of hosts in heavily data base oriented shops, and (3) effective sharing of data by multiple loosely coupled hosts, including dissimilar hosts.

We also expect to see "intelligent" storage hierarchy controllers which work in concert with the DBMS to permit more effective data staging (smaller segments staged with greater predictive accuracy) together with exploitation of data redundancy at various levels of the hierarchy to enhance integrity/recoverability of data. This is an avenue toward practical implementation of very large data bases comprising billions of characters.

Also available will be parallel controllers in which a high level operation such as search and mark is executed on several disks simultaneously to reduce the total search time. This presents a very clear cost/performance tradeoff which would ideally be tunable between the extremes of one processor per disk and one processor per track. The need for indexing structures would be reduced correspondingly, and in the extreme case a simple query could be answered in the time it takes for two revolutions. From another point of view, partial searches performed simultaneously on multiple disks may be an alternative strategy for accessing very large data bases. Such a strategy is relatively expensive but provides faster response characteristics. This technology may not be cost-effective for several years.

7.4.2.2. Ten Year Research Needs. Since data base processors are for dedicated purposes, we would expect in the long run to see research aimed at increasing use of special instruction sets and machine architectures specifically geared to the data base management functions such as searching, sorting, and set intersection.

7.4.3. Distributed Data Bases. The physical distribution of a logically integrated data base over several distinct computing facilities (nodes which are interconnected by some communications facility (link)) is called a distributed data base. Logical integration means that each node has access to the entire data base depending upon DBA imposed restrictions. Ideally, the physical distribution of the data base is transparent to the user. For the purpose of this discussion, the computing facilities consist of processing units with main store, associated secondary storage, and communication capabilities. The nodes may have similar or dissimilar computing facilities.

At each node the software complement consists minimally of an operating system, a data base management system, and communication management. With the exception of the latter, the software components may also be similar or dissimilar.

The distributed data base is realized when the resource sharing concepts are combined with data base technologies. Consequently, the technology facing distributed data bases encompasses not only those issues relevant to both the resource sharing (computer network) and data base areas, but also those issues which result due to the integration of the two areas. Distributed data base issues will be discussed in terms of these categories.

7.4.3.1. Resource Sharing Issues. The resource sharing area encompasses a myriad of operational issues which directly affect the operation of distributed data bases. The configuration and homogeniety of the system determine to a large degree the technology required. Homogeneous systems will naturally require less effort to integrate than heterogeneous systems. The latter require interfaces between hardware configurations, data formats, operating systems and data base management systems. The resource sharing system, with its communication subsystem, must also be examined with respect to its ability to handle large volumes of data while at the same time preserving security and privacy.

In addition to technological issues, issues relating to the effi cient management and usage of the system are also involved. One such issue is the distribution of resource and data to optimize the effi ciency of the resource sharing system. These resources would include application programs, data base programs, and structured data bases. The transparency of the system to the user, an issue concerning ease of usage, is also important because it may ultimately set an upper limit on the level of transparency which may be achieved within a distributed data base system.

7.4.3.2. Data Base Issues. The basic issues today in the distribution of data bases are similar to those which have faced DBMS researchers for the past ten years. Issues concerning centralized versus decentralized data, level of redundancy (multiple copies), privacy, integrity, and security existed long before the advent of distributed data base technology. These issues are, however, further complicated by the autonomous and independent nature of the system. For example, issues such as update, deadlock, reliability, and backup increase considerably in complexity when problems involving multiple copy data files and non-functioning host computers are introduced. Such problems must obviously be taken into account.

7.4.3.3. Integration Issues. Integration issues are the problems which arise when two or more DBMS's and data files are integrated into a distributed system. Among homogeneous data base systems, the level of effort required is small in relation to heterogeneous systems. The basic issue appears to be the development of a control structure. However, the integration of different DBMS's involving different data models, data definition languages, data manipulation languages, and data formats will require a large effort in DBMS translation technologies. Schemes for global control of the system (to achieve transparency, provide translations, record statistics, maintain integrity, etc.) and global addressing techniques (master directories, schemas) are also important issues to be resolved.

The rationale for distributed data bases is the decentralization of the data processing function while sharing data. (The equipment and operating costs for distributed data bases approach those for centralized systems having large configurations of distributed terminals.)

Most distributed systems today are highly customized with much expensive special software. Some customized distributed data base systems currently exist in a prototype form. Examples of commercial applications under development are in the banking,discrete, and continuous process control areas. To our knowledge no distributed data base systems are commercially available today.

7.4.3.4. Five Year Operational Outlook. A trend is already developing toward the implementation of DBMS on small computing systems such as IDMS on the PDP11/45. As a natural outgrowth of this, one can expect to see in the next five years commercially available distributed data base systems within a vendor's product line. These systems are expected to utilize homogeneous data base management systems.

7.4.3.5. Ten Year Research Needs. The goal of data sharing in a multicomputer network intensifies existing problems in data management and introduces a new class of problems. The existing approaches to the single data management system, issues of privacy, integrity, concurrent access, etc., are challenged by the distributed nature of the system. Several areas require additional research prior to long term use of distributed systems. Further integration of nodes within resource sharing systems is required in order to provide a foundation for distributed data base systems. This area would involve the transferability of data, transparency of processes from dissimilar nodes, and the distribution of resources (data and software) to optimize system performance. Synchronization of multiple copies of distributed data must also be investigated. Problems with update, backup, and concurrent access increase in complexity due to the distributed environment. Finally, the capability of the resource sharing and distributing systems to store data and execute programs at any node will depend on the development of DBMS translation technologies, particularly, data query and model translation. These technologies must be developed in order to achieve the integration of DBMS's, which is an initial goal of distributed data sharing systems.

7.5. New Functions

Numerous new developments are occurring in data base technology. Data models, knowledge representation, natural language query systems and others are vital issues. This section summarizes the panel's observations on these developments.

7.5.1. Data Models and Languages. The area of data models and supporting languages is in an era of inventiveness with a number of languages and models either in existence or being proposed. Each of these models has its proponents who are able to point out advantages for their particular model and suggest that these advantages are decisive. However, at present, there is no consensus as to which model is best, nor, considering the complexity of the technical issue and human factors, is there likely to be conclusive support for any one approach within the next five years.

To understand the data base model question, consider the analogous development of automobiles. At the earliest stage there were major technical differences between the competitors. Some had three wheels, some had four; some had steam engines, some had internal combustion engines. Each technology had its proponents who emphasized their technology's advantages and suggested that their technology was best. However, it was not honestly possible to say whether there was a best technology, because some had not reached stable positions on their learning curves. In this situation, customers' personal backgrounds and understandings had great influence on their selection of a particular technology and many technologies were selected and used. It took decades to test out the characteristics of the various technologies and select those which would predominate for various functions.

The field of data base models and languages is in a similar situation today. There exist at least five models with different, sometimes overlapping characteristics: network models, hierarchical models, relational models, binary association models, and set-theoretic models. Although the proponents might not agree, we believe that each model is capable of supporting a corporation's data base system. That is, there seem to be no inherent absolute limitations in what the models can describe. In this sense, they are equivalent. On the other hand, there are much more difficult questions of relative efficiency of each model both with regard to machine efficiency and ease of customer use. There is not yet enough evidence to support the superiority of any particular model. In fact, the user's ease of problem specification using a particular model may now and forever be the most important consideration. Some users will find their problems to be most easily solved in terms of networks, others in terms of hierarchies, others in terms of relations, etc.

7.5.1.1. Five Year Operational Outlook. No data model will magically solve all the manager's data base problems. Fortunately for those users whose problems naturally fit into networks or hierarchies, relatively stable and satisfactory system implementations exist which they can immediately use. On the other hand, the characteristics of more recent models are neither fully developed nor completely understood. At the present time, prototype systems of these more recent models are found to be inefficient with respect to machine utilization, and response time, but present work will certainly improve the efficiency. The differences in efficiency of the fully developed models are likely to be differences in degree rather than differences in kind. In this case, specific user preferences, backgrounds and specific problems are likely to continue to be a major determinant in which system is best for which user.

Research has begun to compare the models in as objective a way as possible with regard to machine efficiency and ease of human use (this latter category includes human factors, studies of both the model itself and the languages which the customer must use to process the model). This research may either show that one model represents the most desirable tradeoff of features or it may conceivably indicate that the specific use to which the system is being put will always be the most important factor in selecting the model. We expect that it will take about five years to gain enough experience with relational systems to determine whether they are to some degree more useful than previous technologies.

If any of the new models demonstrates after 5 - 8 years that it actually offers a significant degree of improvement then manufacturers will provide bridges to the model and, in fact, may build it as a compatible extension of their existing system.

In this case, the user should select the model that presently best fits the background and problem for the near term future. In terms of expected advantages, present proposed new models are at best evolutionary, rather than revolutionary.

7.5.1.2. Ten Year Research Needs. Intensive effort is needed during the next five to ten years to resolve the data model issue. A number of models are now being examined on a limited scale. This research has an important need for information on industries' use of data bases in both batch and interactive modes to help decide which models might be successful and which ones overshoot or undershoot the required mark. The panel suggests increasing the limited effort in this area.

7.5.2. Data Base Semantics

7.5.2.1. Introduction. Users of data base management systems often assign meaning to data which the physical representation does not convey. This meaning or "semantics" specifies the intent of the users, eliminates meaningless operations on the data, or enables the system to make inferences based on the data. There are many ways of dealing with semantics involving different points of view. All approaches are still at the research and prototype level.

One approach attempts to expand the Data Definition Language (DDL) so that additional constraints can be put on some data operations and on some data relationships. In this way, the user can specify requirements, or intended use of the data. These requirements control the operations and the evolution of the data base. For example, if the user knows that "last name, first name, address" identifies people uniquely he may want to enforce this restriction on his data base. This situation is quite independent from the use of a particular key in building an access path. A user may request that "last name, first name, address" uniquely identify a person, while at the same time asking for the construction of an index using only "last name." Much work is done in the syntactic specification of essentially semantic constraints, (Boyce and Chamberlin 1973, Tsichritzis 1975). An interesting idea, for example, is the specification of units for data. In this way $10,000 can be distinguished from 10,000 miles. Such work usually follows the traditional approach in programming languages: giving more meaning to data types by declaring their properties and constraining their operations. For instance, in most languages one cannot multiply numbers and character strings, or append a matrix to a bit string. In the same way a data base system should distinguish the properties of money, for example, and give constraints about the operations on money according to sound accounting practices.

Another approach attempts to capture semantic information in data models. An effort is being made to define a framework in which a user can specify the information requirements of his application, e.g., the infological approach (Sundgren, 1975). In addition, formal models can be used to describe the meaning of data, and to analyze the meaning of queries, e.g., semantic networks (Roussoupoulos and Mylopoulos, 1975), the DIAM II model (Senko, 1975). Techniques for the description and manipulation of knowledge in these models are currently being investigated. In addition, research on the semantic properties of other existing data models is progressing, e.g., semantics of the relational model (Schmid and Swenson, 1975).

7.5.2.2. Five Year Operational Outlook. The trend towards higher level data base languages will continue. Commercial DBMS's will provide some semantic capabilities in the form of statement of constraints and requirements on the data. This development may be associated with some increased operational cost for the application of these semantic requirements on the data base. In addition, users should make an effort to understand the meaning of their operations. A data dictionary is the first step in defining precisely the names of the data items. The specification of additional semantic information will require the thorough understanding of the relationships of data and the meaning of different operations. The system will only provide the tools to describe and use semantic information. The users will have to capture exactly the meaning and purpose of their applications in order to use the tools properly.

7.5.2.3. Ten Year Research Needs. The research approaches discussed will evolve and they will eventually relate to each other. Systems with increased semantic knowledge of their environment will become realistic. Hopefully, we will have:

(1) Easy to understand and powerful model(s) to describe semantic information.

(2) A complete set of DDL facilities to capture the semantic information described in a model.

(3) A good way of mapping the models to schemas using DDL facilities.

(4) A system which can use semantic information encoded in the schema properly and without excessive overhead.

REFERENCES

1. R. Boyce and D. Chamberlin, "Using a Structured English Query Language as a Data Definition Facility," IBM San Jose, Report RJ1318, December 1973.

2. N. Roussopoulos and J. Mylopoulos, "Using Semantic Networks for Data Base Management," Proc. 1975, Boston Conference on Very Large Data Bases.

3. H. Schmid and J. R. Swenson, "On the Semantics of the Relational Data Model," Proc. ACM-SIGMOD Conference, San Jose, California, May, 1975.

4. M. Senko, "The DDL in the Context of Multilevel Structured Description: DIAM II with FORAL" In Data Base Description, Donque and Nijssen, North Holland, 1975.

5. B. Sundgren, Theory of Data Bases, Prentice-Hall, December, 1975.

6. D. Tsichritzis, "Features for a Conceptual Schema." CSRG Report 56, 1975, University of Toronto.

7.5.3. Relational Inferences in Data Base Management Systems. The idea of inferences in data base management systems deals with the development of explicit data from implicit data, the making use of data base semantics. An example of an inference is the derivation that a particular individual is a GRANDFATHER of an individual given only the FATHER relation and a general rule that "the father of the father or the father of the mother is the grandfather." Most systems have some inferential capability. Such a capability is generally achieved by either contiguity or by data structure. An example of contiguity is where one has a record with an individual and offsprings. To find the sibling of an individual, one finds the parent's record, and the set of individuals in contiguous positions in the offspring portion of the record are the siblings. An example of an inference through data structure is one in which there is a link from an individual's record to the parent record. As in contiguity, one finds the parent record, but then finds the link in the parent's record to find the grandparent of the individual.

Some research systems have more sophisticated inference capabilities. Two instances of these are LUNAR being developed by Woods and Minker's MRPPS. These systems generally have three approaches to inferences: (1) Built-in inference procedures which permit a small number of general rules of inference to be used, (2) Inferences using the predicate calculus and theorem proving techniques which can handle an unlimited number of general rules, and (3) Inference procedures generated through a procedural language.

To achieve a general inference capability as just described, current generation DBMS's will require a general problem solving capability. Current DBMS's cannot be considered to have a general inference capability.

7.5.3.1. Five Year Operational Outlook. Work on inference development will be performed primarily at universities and in some research centers. To make such systems practical, the following will have to be developed:

A. Heuristic techniques that guide the search.

B. Use of real world knowledge in the form of semantic information will have to be used to control the search. The manner in which one uses and represents semantic information must be established.

C. An effective system will require interactive response with the user.

D. The amount of syntactic vs. semantic information needed to control the search must be determined.

E. The effectiveness of the techniques for large scale vs. small systems must be established.

From what is now known, we can attain important insights in the next five years. However, we will not resolve all problems by that time.

7.5.3.2. Ten Year Research Needs. In a ten-year period we may see relational systems having an advanced inference capability. The DBA will have to determine the degree to which data should be explicit or implicit. Once determined he will be able to specify the general rules and other information required to make data in implicit form explicit. With such a tool developed, the user will not need to specify how to develop a new relational form described as some combination of given relations. If the general rules are already in the system, he need only supply the name of the new relation and the system will develop it automatically.

During the next ten years, this technology will have only a slight impact upon business and management.

REFERENCES

1. Minker, J., "Performing Inferences Over Relational Data Bases," Proc. of the 1975 SIGMOD Conference, May 1975, San Francisco, California.

2. Minker, J., "Set Operations and Inferences Over Relational Data Bases," Fourth Texas Conference on Information Science, November 17-18, 1975, Austin, Texas.

3. Woods, W. A., Kaplan, R. M. and Nash-Webber, B., "The Lunar National Language Information System," Final Report, Bolt Beranek & Newman, Inc., Cambridge, Mass., 18 June 1972 (BBN) Report No. 2378).

7.5.4. Natural Language Query Systems

7.5.4.1. Introduction. Although work in natural language analysis has matured over the years, the efforts are not yet adequate to handle unrestricted natural language. There appears to be little likelihood that a system will be developed to handle unrestricted natural language regardless of the amount of time and energy expended upon the effort. However, one need not have full natural language to have an adequate query system. Most queries specified by users are simple in nature since it is difficult to conceptualize complex statements. Thus, many of the problems that arise in natural language analysis may be avoided.

Current query languages use English words in simple forms which appear to be English-like in nature. The range of work in natural language varies from standard sentence template forms through the Chomsky language hierarchy (deterministic, context-free, context-sensitive, and unrestricted), to transformational language, case grammars, frames and procedural languages. The more complex the approach the closer one approaches "natural language," and the more processing required.

It is currently unclear whether or not it is even desirable to provide a near natural language capability for a DBMS. Highly stylized languages based on simple models of natural language may be all that is required. Systems which provide a dialogue capability for the user have been developed and used in military applications. They have not been overwhelmingly successful.

7.5.4.2. Five Year Operational Outlook. Studies are needed to determine the utility of natural-like language to highly structured languages for DBMS.

The work by Woods on LUNAR has shown that within a particular domain we can develop a relatively sophisticated natural-like language approach. Progress going beyond what Woods has achieved will be very difficult and of questionable utility.

We should experiment with an interactive natural-like language approach employing a dialogue between the user and the system and establish the problems associated with this approach. Such experimentation would have great utility if performed on a large scale data base.

We should also experiment with the use of semantic information based upon the domain of application and provided by semantic networks. This work would establish the ease in which one can go from one domain of application to another, and the effect of changing domains on the complexity of the data base query language. The manner in which a query language is integrated as part of a general data base language should also be established.

7.5.4.3. Ten Year Research Needs. During the second five year period, a more natural-like query and data base language capability should exist. Management can expect to be able to use a more natural manner of addressing queries and commands for Data Base Management Systems.

REFERENCES

1. Minker, J., "Information Storage and Retrieval - A Survey and Functional Description," University of Maryland TR-369, April 1975.

2. Woods, W. A., Kaplan, R. M. and Nash-Webber, B., "The Lunar Sciences National Language Information System," Final Report, Bolt Beranek & Newman, Inc., Cambridge, Mass., 18 June 1972 (BBN Report No. 2378).

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

4. Thompson, F. B., et. al., "REL: A Rapidly Extensible Language System," Proc. 24th National ACM Conf., Aug. 1969, 399-417.

7.6. Miscellaneous

7.6.1. The Effect of Standards Upon Evolving Technology. The Panel discussed the effect of standards upon data base technology. The difficult question of "when" to standardize was addressed in some detail by the Standardization working panel of this workshop.

Premature standardization will certainly impede the evolution of a technology and more seriously could prevent users from keeping pace with progress. On the other hand, from standards we find the foundation from which to launch the next stage of evolution. Although this topic was seriously discussed the Panel did not reach any conclusions in this area. Each proposed standard must be considered in its very detailed specifications before any position can be taken on its merits in terms of what it will accomplish and cost. We must ascertain that any standardization that occurs will not adversely affect evolution to new data base models, to more powerful data manipulation languages and more flexible use of data base systems.

7.6.2. Research and Financing. Research in data base technology has a very high cost. More than other research, it involves both (1) detailed exploration of how to model and measure data base systems and, (2) experimental analysis and verification of the results. The panel classified research appropriate for a doctoral candidate as that which can be accomplished within two or three years and has a definite success product and measure of creation or inventiveness. Much of the existing work needed in data base system development involves measuring existing DBMS systems, developing models of existing systems and seeing how they differ and compare with proposed theoretical models. Such work either fails the time requirement or the appropriateness for PhD work. Though this type of work needs to be promoted, we do not see ways in which it can be done.

We also note the nationwide decrease in the total amount of research in computer science with the demise of the research activity in several of the major computer manufacturers. The amount of research going on is not as great as could be because of the smaller number of PhD students which are being graduated each year by Computer Science Departments.

The panel recognizes the need for more research in many of the areas described above. As one of the panel members put it, "industry is our laboratory." Computer science research groups need information on existing data bases and, in fact, need experimental data bases which can be analyzed. They need dialogue with users to learn the present and projected uses of data base management systems and they need more expertise within their own ranks from people who have used systems and can reflect on results of such uses.

The panel proposes that an intensive effort be made to encourage cooperative data base research between research groups and industrial users which have data base systems that can be modeled and measured. Research groups can use non-sensitive data base systems as laboratories for experimenting with new types of data base systems. Industry, on the other hand, must realize that laboratory successes require verification _in vivo_ and that in the end their data base systems will perform much better if they participate in joint research activity.

8. Background

8.1 Introduction

The reader will have a greater appreciation of the reports in these proceedings if he understands the background activities that lead to the workshop.

Richard Canning and Jack Minker, acting as liaison between NBS and ACM, brought the idea of a workshop on data base systems to Seymour Jeffery at NBS. The topic of data base technology fitted well into the series of joint NBS/ACM workshops on major computer issues inaugurated in 1972 by Dr. Ruth Davis, Director of the Institute for Computer Sciences and Technology, and Walter Carlson, then President of the ACM. After some discussion of the purpose and structure of the workshop, NBS and ACM established a planning group to develop the workshop format, set a time-table, and the working panel subjects. Richard Canning agreed to chair the workshop. The planning group selected the panel chairmen, who also became members of the planning group.

The enlarged planning group determined the subject matter to be covered by each panel and developed a set of questions for the working panels. Once the questions were set, each panel chairman selected the members of his panel. The questions were distributed to the panel members and they were asked to prepare answers for circulation to the other panel members prior to the workshop.

On October 29, 1975, the workshop began two and a half days of intensive effort. The approximately 80 participants met in a plenary session to hear the keynote speaker, Daniel Magraw. By mid-morning the workshop had received its instructions to develop the information needed by a manager considering the use of data base technology. From that point until the closing plenary session, each working panel met separately to collect, discuss, analyze, and compile the information seen here.

During its closing session, the workshop participants heard each of the working panel chairmen present his panel's report. Each panel report was followed by a question period. After all the reports had been discussed, the workshop turned to a general discussion of "what next?"

8.2 Organization of the working panels

Though each working panel approached its task in a slightly different style, common to all were the responsibilities of the panel chairman and the recorder. The chairman guided and paced the discussion and ultimately had the task to prepare the panel's report.

The recorder was a member of his panel selected to maintain the panel's minutes. The recorder kept notes on a flipchart pad so that all could see how the minutes were recorded. Each completed page was displayed during the ensuing discussions to help focus the points being made. Periodically, the displayed pages were collected, typed, duplicated and distributed to panel members. Thus, by close of the workshop, each member had a complete set of mutually agreed upon notes. From these notes, the chairman prepared the panel report. Time prevented the circulation of the panel report at the workshop but the reports were edited, polished and (in some cases) circulated to the panel members several weeks after the workshop but prior to submission to the proceedings editor.

The working panel reports were compiled with other information from the workshop to make the proceedings more useful and readable. The primary reading audience, of course, is managers facing a decision about data base systems. The secondary audience is the several technical disciplines that assist and support managers.

8.3 Conclusion

The five vantage points (auditing, government regulation, evolving technology, standards, and user experience) used to survey data base systems contribute to broadening each of these five viewpoints. By reminding ourselves of the manager facing data base decisions and the importance of his needs, the proceedings may foster better understanding between the users and providers of data base technology. As a concrete record, these proceedings provide a foundation for future efforts.

Appendix A

Criteria to be Applied in the Standardization of a Programming Language

ISO

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION

ISO/TC 97/SC 5

NOVEMBER 1967

CRITERIA TO BE APPLIED IN THE STANDARDIZATION OF A PROGRAMMING LANGUAGE

INTRODUCTION

The purpose of this document is to present criteria to be applied in the standardization of programming languages.

There are two types of occasion when criteria should apply:

(a) First when a language is considered as a candidate for standardization*, see item 1;

(b) Second when a document or documents describing a language are considered as draft proposals, see items 2 to 4.

The criteria for a candidate pertain to the attributes of a language such as its need, utility and general acceptance. The criteria for the documentation pertain to its style and content.

It is recognized that the standardization process must be evolutionary and must encourage and not impede developments in computer applications and languages. Therefore, these criteria are designed to facilitate the standardization of currently used programming languages, to provide for the further development of existing languages, and to encourage the consideration of emerging languages.

It is further recognized that the field of language specification has, as yet, not produced a universally acceptable methodology. This is an urgent necessity. It is expected that further work in this area will be forthcoming. In the meantime, the criteria herein presented emphasize the use of any existing methodologies which will serve the purpose of generating an acceptable standard.

---

*A candidate language for standardization is a language for which the Committee, currently known as ISO/TC 97/SC 5, has agreed to process an ISO Recommendation.

The design and implementation of programming languages is a complex and relatively new art, and there is, as yet, little experience in bringing programming languages within the scope of standardization activities. Thus, while the present list of criteria is complete in the current state of the art, developments in the techniques of language specification will require a continuing revaluation of the criteria themselves.

1 - BASIC CRITERIA

Before a language is considered as a candidate for standardization, acceptable bodies should be identified who will be responsible for:

(a) submission of the proposal,
(b) modifications in the light of ISO requests,
(c) maintenance.

2 - CRITERIA FOR SUITABILITY OF PROGRAMMING LANGUAGE STANDARDIZATION

The following requirements must be met for a language to be accepted as a standard:

(a) A substantial number of prospective users of the standard language must exist in the area of application.
(b) The language must accommodate a substantial portion of the problems confronting the intended users.
(c) The language should be compatible with those standards, recommendations and accepted practices which are considered applicable. Deviations and discrepancies must be justified.
(d) The language must be such that a processor for the language can be implemented with hardware and software facilities generally available to the intended users.

3 - CRITERIA FOR DRAFTING AN ISO RECOMMENDATION

In drafting an ISO Recommendation the following criteria apply:

(a) A Draft Proposal should, and a Draft Recommendation must, be prepared in the format and in the style required by the Guide for the Presentation of ISO Recommendations. Devices such as a Table of Contents and an Index are recommended where they will facilitate the use of the document.
(b) The definition of the language must be clear, precise and self-consistent.

The rigorous use, where appropriate, of well-defined metalanguages, diagrams, etc., is preferred, but concise natural language may be acceptable. In some cases processing algorithms may be required for adequate definition. Any combination of techniques may be used to enhance clarity of definition. Usage of these techniques must be compatible with related ISO Standards and Recommendations in the field.

(c) The description of the language must be such that any program written in the language is capable of one and only one interpretation according to the proposal. In that regard, elements having an interpretation which is indefinite must be identified.

(d) Design considerations, hardware and media representations, specific criteria, justifications, and historical information are generally preferred as appendices rather than as part of the Recommendation.

4 - CRITERIA FOR CONSIDERATION OF LANGUAGE MERIT

The procedures for standardization of programming languages do not impose requirements on the intrinsic characteristics of a language and do not stipulate the manner in which a language is recognized as being a programming language. Such prescriptions are not to be inferred from this specification.

Nevertheless, it is difficult to imagine that evaluation for standardization would occur without some consideration of intrinsic features. While this document does not prescribe criteria for such characteristics, nor weights to be attached, nor points of application, it is clear that criteria such as the following apply at least informally.

(a) It should not be needlessly difficult for the intended user to learn the language.

(b) It should be natural to write programs in the language which are easily understandable to the intended users of the language.

(c) The language should have no arbitrary limitations or exceptions in its rules. Since this objective may be compromised by other requirements, any limitations should be clearly justified with respect to such requirements, e.g., learning ease, processing efficiency, available capacity.

(d) The language should provide the intended user with appropriate access to facilities for effective communication with the environment.

(e) It is desirable that the language should lend itself to the construction of programs which may be subdivided and the resulting pieces separately written and tested.

(f) There may be standard subsets, but unnecessary proliferation is to be avoided.

Appendix B

The Study of Data Base Management Systems With Bibliography

The following paper was submitted to the participants of the Workshop and provides a good overview of the material available in the area of data base management systems. The paper's introductory material and annotations guide the reader through a burgeoning thicket of articles addressing this increasingly important subject. We are indebted to Drs. Chester M. Smith, Jr. and Barry R. Munson for permitting us to reproduce it in its entirety as an Appendix.

The Study of Data Base Management Systems with Bibliography

by

Barry R. Munson

and

Chester M. Smith, Jr.

revised to
January, 1976

Computation Center
The Pennsylvania State University
University Park, Pennsylvania

Computer Science Department
The Pennsylvania State University
University Park, Pennsylvania

ABSTRACT

The purpose of this paper is to provide a basis for the study of data base management systems. While presenting somewhat of an overview, it is not intended that the overview be comprehensive in itself. Rather, it is to serve as a skeleton for a further comprehensive study by the reader. The extensive bibliography, which is partially annotated within the text, should be most helpful in this regard.

INTRODUCTION

The study of data base management systems (DBMS) encompasses a field of inquiry which is nearly as large as computer science itself. Highly relevant areas include information retrieval, operating systems, list processing techniques, management information systems, real-time systems, management of data processing, file organizations, hardware design, searching, sorting, integrity, and security.

Much of the current literature on DBMS has been devoted to user-system interfaces, objectives, requirements, and approaches. Since the field is in its infancy (comparable to when operating systems were first allowing multi-programming), there has also been a large amount published on the desirability of DBMS.

There is general agreement on the major objectives of a DBMS:

1. The data base should be shared (by definition).
2. A high level of data independence, software independence, and hardware independence should be accommodated.
3. Most (if not all) data redundancy should be eliminated.

4. Data relations must be effectively handled.

5. Security should be provided.

6. A high degree of data integrity must be assured.

7. The system must be cost-effective.

There is also general agreement that the following entities constitute the major functional areas which are needed for an effective DBMS.

1. The Data Manipulation Language (DML).

2. The Data Definition Language (DDL).

3. The Data Dictionary/Directory (DD/D).

4. The Data Base Administrator (DBA).

Most authors feel that as a further objective the DML, DDL, and DD/D should be independent of each other and that their structure should be independent of the data base itself.

In contrast to these areas of general agreement, there has been much debate on <u>how</u> these various objectives are to be met, the extent to which they are to be met due to trade-off considerations, and on the design criteria for the functional areas.

OVERVIEWS

The best way to begin the study of any field is with a concise overview of the field. There is no shortage of such articles on DBMS. Due to the vastness of the field, however, the articles are necessarily shallow and are generally restatements of each other.

Suggested readings include the following:

1. Plagman* is the principle author of a series of portfolios in Auerbach's Data Processing Manual which describe DBMS concepts, planning, implementation, and administration.

2. COBOL data base facilities specified by the CODASYL Data Base Task Group are also described in a series of Auerbach* Data Processing Manual portfolios and are also very relevant to an overview of current DBMS thinking.

3. Canning,* as editor of EDP Analyzer, has devoted a number of issues to overviews of DBMS. Though each cover a certain specific area, they also provide the reader with a good general overview.

4. Bachman's* 1973 ACM Turing Award Lecture was presented in the Communication's of the ACM (November, 1973). The overview describes the

---

* Throughout this paper, "*" is used to denote references to the bibliography at the conclusion of the paper. Though unnumbered, the text provides enough information to allow the reader to easily find the reference work.

programmer as a navigator in the data base structure.

5. Lupien* presented an overview of the GUIDE-SHARE DBMS Requirements at GUIDE (May, 1971).

6. Price* was the panel chairman at a GUIDE (November, 1972) session which discussed managerial considerations for DBMS. The paper developed from this session is an excellent overview.

7. Patterson* presented "Requirements for a Generalized Data Base Management System," at the Fall Joint Computer Conference (1971).

8. Whitney* presented "Fourth Generation Data Management Systems" at the NCC (1973).

Many other overviews are included in the bibliography and their titles generally make this apparent.

MAJOR WORKS

Having obtained an overview of DBMS through short articles, it is suggested that the study continue with some of the major works which have been written. In particular, the Joint GUIDE-SHARE DBMS Requirements* and the CODASYL Data Base Task Group* report are referenced by most authors in the field. Another commonly referenced work is the CODASYL Systems Committee* report on feature analysis of generalized DBMS.

Suggested readings include the following:

1. The Joint GUIDE-SHARE Data Base Requirements Group* report describes their view of long-range requirements for DBMS. The emphasis is on "long-range" and on "requirements." Not all of the listed requirements are "necessarily realizable on current hardware and software systems," as is stated in their introduction. Furthermore, the report is confined to requirements as opposed to the details of how these requirements may be achieved. However, the report is an excellent summary of the ideals which are to be sought in a DBMS.

2. In contrast to the GUIDE-SHARE report, the CODASYL Data Base Task Group* report proposes a currently achievable common approach to DBMS. It describes in detail its proposed DDL for the programmer (used in the sub-schema) and a DDL for describing the data base (used by the Data Base Administrator to develop a schema).

   It also details a proposed DML for use in COBOL. Many commercially available data base management systems have already been patterned after these recommendations. Noteworthy by its absence is IBM, which has many reservations about the desirability of the suggestions for a common approach.

3. Martin's <u>Principles of Data Management</u> is excellent in its comprehensive coverage of the major sub-areas of interest in the study of DBMS. The text includes explanations of a range of logical organization structures, physical organization structures, the CODASYL Data Description Language, IBM's Data Language/I, relational data base approaches, and a comprehensive selection of related objectives and techniques.

4. The CODASYL Systems Committee* report on <u>Feature Analysis of Generalized Data Base Management Systems</u> defines the features offered in present day systems. Eight commercially available systems and the Data Base Task Group proposals are described in relation to each of the ten features described. In addition COBOL is considered as a basis for further development. An introduction written by the CODASYL Systems Committee* may be found in the <u>Communications of the ACM</u> (May,1971).

More recently the CODASYL Systems Committee presented a summary of a follow-up report at

ACM'75. This report describes in detail the criteria which should be considered in the Selection and Acquisition of Data Base Management Systems.

5. Cagan's* Data Management Systems provides a very well written elementary overview of DBMS. The book is a first choice for the non-technical or uninitiated reader.

6. Tou's* Information Systems is a collection of papers submitted at the International Symposium on Computer and Information Sciences (COINS) in 1972. The first seven papers (approximately one-third of the book) are devoted to DBMS. Particularly impressive is Everest's* paper on The Objectives of Database Management, which was taken from his Ph.D. dissertation.

7. Knuth,* volume 3, on Sorting and Searching must be included for completeness. Though the volume is not particularly oriented toward DBMS, there is an abundance of material which is directly related to DBMS implementation. His well-known works and expertise need no introduction.

8. The Quarterly Bibliography of Computers and Data Processing is an excellent source of material for the study of DBMS. A large variety of

periodicals and books are included with brief annotations following each. Much of the material used by the authors was located through the use of this source.

THE DEBATE

The debate on DBMS has centered on comparisons between the GUIDE-SHARE report and the CODASYL Data Base Task Group (DBTG) report. In addition, IBM, as a member of the DBTG, submitted a minority report stating its objections to incorporating the proposed DML and DDL into the COBOL Journal of Development. Therefore, to a large extent, the sides have been IBM and its joint user groups versus the DBTG.

A direct comparison between the GUIDE-SHARE report and the DBTG report is not completely appropriate since the former is an "ivory tower" approach while the latter is a currently realizable and completely feasible approach. The primary arguments against the DBTG proposal stem from the degree to which the proposal falls short of the data independence envisioned in the GUIDE-SHARE report.

Suggested readings include the following:

1. Canning's* <u>EDP Analyzer</u> article on the subject presents a very good summary of the main points of

each of the two committee reports and of the arguments for and against the DBTG proposal.

2. Engles,* as the IBM representative on the DBTG, described his arguments at the ACM-SIGFIDET Workshop in 1971.

3. IBM's* objections to the DBTG were presented to the CODASYL Programming Language Committee (the parent committee to the DBTG) in 1973.

4. Jardine* presented "A Critical Analysis of Data Base Requirements"at GUIDE (1972). He was a consultant to the GUIDE-SHARE group, and in the paper he presents some of his reservations on the DBTG proposal. The paper also includes a very good data independence discussion which he expanded upon at SHARE (1973).

5. Collmeyer* proposed an alternative to the CODASYL DML at the ACM-SIGFIDET Workshop in 1972.

6. Tani* presented a comparison of the DBMS reports at SHARE (1972).

7. Parsons* was listed as the primary author of a Computer Journal (May,1974) article which noted the problems with boolean operations using the DBTG DML.

The Data Dictionary/Directory (DD/D) in a DBMS is a central location for data descriptions maintained by the Data Base Administrator using the DDL. The dictionary contains source data definitions including descriptive text for users, while the directory contains object data definitions which direct the computer system to the physical data.

A Data Dictionary/Directory System (DD/DS) can also exist independent of a DBMS, and several such systems are commercially available. As Uhrowczik* points out, "Although the objectives of a DD/DS are similar to the often-cited objectives of a DBMS, to a certain degree these can be achieved even outside of a DBMS environment by means of DD/DS. However, the combination of a DD/DS and DBMS can achieve these objectives to a much higher degree than can either by itself."

Suggested readings include the following:

1. Uhrowczik's* article in the IBM Systems Journal presents a very well written view of the DD/D concept. The capabilities, objectives, and contents of a DD/D are described along with relational descriptions and an implementation example. Many diagrams and charts are also

included.

2. Cahill* describes a DD/D for building a common MIS data base in a much referenced Journal of Systems Management article (November, 1970).

3. Canning* gives a good overview of specific DD/DS packages which are commercially available in EDP Analyzer (November, 1974).

DATABASE ADMINISTRATION

The Data Base Administrator (DBA) function consists of the person or persons with the responsibility of maintaining the integrity of the data base and for its efficient organization. The DBA is also responsible for defining the rules and data descriptions for its use. The function is not to be confused with the Data Base Manager which is the software and hardware of the DBMS; and which is commonly thought of as being the DBMS. In the larger sense, however, it is generally agreed that a DBMS is not complete or viable without the central control functions of the DBA.

Suggested readings include the following:

1. The SHARE Data Base Administration Committee* report (June, 1974) presents a very comprehensive view of the requirements, duties, and capabilities needed by the DBA. The suggestions for staffing

this function are particularly informative.

2. Schneider* presented an excellent overview of the DBA at GUIDE (1972) in a short paper which is more than just a "glossing over" of the subject.

3. Uhrbach* also presented a paper at the same GUIDE Conference in which he suggested the experience and educational background which is needed by the DBA.

FILE ORGANIZATION

Perhaps the function of the DBA which requires the greatest amount of technical expertise is in the area of organizing (and reorganizing) the data base. Decisions must be made on the most efficient organization, not only for one application or department, but also by taking into consideration the organization as a whole. Reorganizations must be made when the efficiency deteriorates over time; the extent and timing of which must be made using cost-benefit analysis.

The three following issues of the Communications of the ACM are recommended readings:

1. Bachman,* who received the 1973 Turing Award as was previously noted, presented an overview of storage structures in the special 25th Anniversary

Issue of the CACM (July,1972).

2. Shneiderman* wrote an article on "Optimum Data Base Reorganization Points" in June, 1973.

3. Cardenas* wrote on the evaluation and selection of file organization in September, 1973.

SELECTION OF A SPECIFIC SYSTEM

For most firms, the use of a commercial DBMS package is a wiser choice than in-house development. The determination of which commercial DBMS package to select, however, is an extremely critical and complex managerial decision. Canning,* in the February, 1974 issue of EDP Analyzer, lists the major families of such systems as follows:

1. IBM Families.

BOMP	Bill of Material Processor
DBOMP	DB Org. and Maint. Processor
MRP	Materials Records Processor
IMS	Information Management System
GIS	Generalized Information System

2. CODASYL Families.

IDS	Honeywell - Integrated Data Store
CODASYL (DBTG)	
DMS 1100	Univac
DMS	Xerox

TOTAL **	Cincom	(** possibly in this family)

3. File Management Families.

MARK IV	Informatics, Inc.
ASI-ST	Application Software, Inc.

4. Inverted File Families.

SYSTEM 2000	MRI Systems Corporation; NCR
ADABAS	Software Ag (West Germany)
IFAM/II	Computer Corporation of America
METABASE	GTE Information Systems

Analyses of these various systems may be found in the following references:

1. Canning* reviews the "competitive ideologies" of the families of DBMS presented above in "The Current Status of Data Management" issue of <u>EDP Analyzer</u> (February, 1974) and gives a basic overview of each.
2. Canning* also devoted a following issue (October, 1974) to systems based on the DBTG proposal.
3. The CODASYL Systems Committee* reports describe the features and criteria which should be considered in choosing from among the various commercial systems. (See MAJOR WORKS section.)

4. Fong, Collica, and Marron describe features and user experiences for six DBMS packages in a National Bureau of Standards report, (November, 1975). The report presents a concise overall analysis of each system aimed at aiding the data base decision maker.

5. Falor* compiled a survey of DBMS software packages in Modern Data, (May, 1971).

6. Numerous presentations on the IBM families, particulary IMS, have been given at GUIDE and SHARE conferences.

RELATIONAL DATA BASE SYSTEMS

The greatest hope for achieving the level of data independence envisioned by the GUIDE-SHARE report lies in the relational model proposed by Codd* in 1970. Since that time much research has been done in this area, yet some practical problems still remain.

The approach is based on relational algebra and relational calculus. Basically the model consists of a number of named relations which associate fields within the data base to form a set. For example, the relation "PART" could "return" a set of triples consisting of part number, supplier, and quantity. The principal advantage of this approach lies in the fact that the user of the relation need not be concerned with the structure of the data base.

Suggested readings include the following:

1. Codd's* pioneering work in the Communications of the ACM (June, 1970) is required reading. The well written article contains a great deal of depth and insight.

2. Jervis* and Parker presented "An Approach for a Working Relational Data System" at the ACM-SIGFIDET Workshop in 1972.

3. Date's* paper in Tou's* book is an excellent tutorial on the subject. Comparisons are given of the relational model approach versus the traditional hierarchical and network approaches.

SUMMARY

It is hoped that the previous overview and reading selections have provided the dimensions of the current literature on DBMS. Of course a comprehensive view of this vast field can only be attained by actual study of the literature. To facilitate this study, a bibliography follows in which much of what has been published on the subject in this decade is included. It is confined to the 1970's in the belief that all of the important concepts of DBMS may be found in works published within the last six years.

BIBLIOGRAPHY

Altman, Astrahan, and Fehder.
"Specifications in a Data Independent Accessing Model."
ACM-SIGFIDET Workshop, (1972), pp. 363-381.

Altshuler, Gene and Plagman, Bernard.
"User/System Interface Within the Context of an Integrated Corporate Data Base."
National Computer Conference, (1974), pp. 27-33.

Applied Computer Research.
Quarterly Bibliography of Computers and Data Processing.
Phoenix: Applied Computer Research, Inc., (1970-).

Astrahan, Altman and Fehder.
"Concepts of a Data Independent Accessing Model."
ACM-SIGFIDET Workshop, (1972), pp. 349-362.

Auerbach.
"COBOL Data Base Facilities - Part I: Introduction."
Data Processing Manual, (portfolio #03-07-01, 1974).
Philadelphia: Auerbach, 12 pp.

Auerbach.
"COBOL Data Base Facilities - Part II: Applications Programming."
Data Processing Manual, (portfolio #03-07-02, 1974),
Philadelphia: Auerbach, 11 pp.

Auerbach.
"COBOL Data Base Facilities - Part III."
Data Processing Manual, (portfolio #03-07-03, 1974).
Philadelphia: Auerbach, 15 pp.

Bachman, C. W.
"The Evolution of Storage Structures."
Communications of the ACM, vol. 15, (July, 1972), pp. 628-634.

Bachman, C. W.
"The Programmer as Navigator."
Communications of the ACM, vol. 16, (Nov., 1973), pp. 653-658.

Bachman, C. W. and Bouvard, J.
"Architecture Definition Technique."
ACM-SIGFIDET Workshop, (1972), pp. 257-306.

Blanchard, J. S.
"We Bet Our Company on Data Base Management."
Datamation, vol. 20, (Sept., 1974), pp. 61-65.

Bontempo, Charles J.
"Data Resource Management."
Data Management, vol. 11, (Feb., 1973), pp. 31-37.

Bridgeman, Keene A.
"Use of Shared Data Files."
Data Management, vol. 11, (Sept., 1973), pp. 78+.

Brunn, Roy.
"On the Technical Side: Ingredients of a Data Base."
Information Systems, vol. 20, (Dec., 1973), pp. 32-34.

Brunn, Roy.
"Trends in Database Management."
Information Systems, vol. 21, (June, 1974), pp. 57-59.

Bryan, N.
"Defining the Requirements for a Data Management System."
Proceedings of GUIDE 34, (May, 1972), pp. 412-419.

Cagan, Carl.
Data Management Systems.
Los Angeles: Melville Publishing Company, (1973), 140 pp.

Cahill, John J.
"A Dictionary/Directory for Building a Common MIS Data Base."
Journal of Systems Management, vol. 21, (Nov., 1970), pp. 23-29.

Canning, Richard G. (ed.)
"The Cautious Path to a Data Base."
EDP Analyzer, vol. 11, (June, 1973), 13 pp.

Canning, Richard G. (ed.)
"Creating the Corporate Data Base."
EDP Analyzer, vol. 8, (Feb., 1970), 14 pp.

Canning, Richard G. (ed.)
"The Current Status of Data Management."
EDP Analyzer, vol. 12, (Feb., 1974), 14 pp.

Canning, Richard G. (ed.)
"The Data Administrator Function."
EDP Analyzer, vol. 10, (Nov., 1972), 14 pp.

Canning, Richard G. (ed.)
"The Data Dictignary/Directory Function."
EDP Analyzer, vol. 12, (Nov., 1974), 13 pp.

Canning, Richard G. (ed.)
"Data Security in the CDB."
EDP Analyzer, vol. 8, (May, 1970), 14 pp.

Canning, Richard G. (ed.)
"The Debate on Data Base Management."
EDP Analyzer, vol. 10, (Mar., 1972), 16 pp.

Canning, Richard G. (ed.)
"Long-Term Retention of Data."
EDP Analyzer, vol. 11, (July, 1973), 13 pp.

Canning, Richard G. (ed.)
"Organizing the Corporate Data Base."
EDP Analyzer, vol. 8, (Mar., 1970), 14 pp.

Canning, Richard G. (ed.)
"Problem Areas in Data Management."
EDP Analyzer, vol. 12, (Mar., 1974), 13 pp.

Canning, Richard G. (ed.)
"Processing the Corporate Data Base."
EDP Analyzer, vol. 8, (Apr., 1970), 14 pp.

Canning, Richard G. (ed.)
"Trends in Data Management."
EDP Analyzer, vol. 9, (May, 1971), 13 pp.

Canning, Richard G. (ed.)
"Using Shared Data Files."
EDP Analyzer, vol. 8, (Nov., 1970), 14 pp.

Canning, Richard G. (ed.)
"What's Happening with CODASYL-Type DBMS?"
EDP Analyzer, vol. 12, (Oct., 1974), 14pp.

Capraro, Gerard T. and Berra, P. Bruce.
"A Data Base Management Specification Model."
National Computer Conference, (1974), pp. 53-56.

Cardenas, Alfonso F.
"Evaluation and Selection of File Organization - A Model and System."
Communications of the ACM, vol. 16, (Sept., 1973), pp. 540-548.

Charles, William R.
"Some Systems Shouldn't Use Chained File Techniques."
Data Management, vol. 11, (Sept., 1973), pp. 33-37.

CODASYL Data Base Task Group.
Data Base Task Group Report.
New York: Association for Computing Machinery, (Apr., 1971), 269 pp.

CODASYL Data Description Language Committee.
Data Description Language - Journal of Development.
Washington: National Bureau of Standards Handbook 113 (1973), 136 pp.

CODASYL Fortran Data Base Manipulation Language Committee.
CODASYL Fortran Data Base Facility.
University Park, Pa: CODASYL, (1976), 125 pp.

CODASYL Programming Language Committee.
CODASYL COBOL Journal of Development.
Toronto: CODASYL, (1976).

CODASYL Systems Committee.
Feature Analysis of Generalized Data Base Management Systems.
New York: Association for Computing Machinery, (1971) 320 pp.

CODASYL Systems Committee.
"Introduction to: Feature Analysis of Generalized Data Base Management Systems."
Communications of the ACM, vol. 14, (May, 1971), pp. 308-318.

CODASYL Systems Committee.
Selection and Acquisition of Data Base Management Systems. (Summary presented at ACM′75.)
Monroevillle, Pa: CODASYL, 262 pp.

Codd, E. F.
"A Relational Model of Data for Large Shared Data Banks."
Communications of the ACM, vol. 13, (June, 1970), pp. 377-385.

Codd, E. F.
"Understanding Relations."
FDT Bulletin of ACM-SIGFIDET, vol. 5, (June, 1973), pp. 13-43.

Collard, Albert F.
"A Data Dictionary/Directory."
Journal of Systems Management, vol. 13, (June, 1974), pp. 22-25.

Collen, Morris F. (ed.)
Hospital Computer Systems.
New York: John Wiley & Sons, (1974).

Collmeyer, A. J. and Shemer, J. E.
"Analysis of Retrieval Performance for Selected File Organization Techniques."
Fall Joint Computer Conference, (1970), pp. 201-210.

Collmeyer, A. J.
"Database Management in a Multi-Access Environment."
Computer, vol. 4, (Nov./Dec., 1971), pp. 36-46.

Collmeyer, A. J.
"Implications of Data Independence on the Architecture of Database Management Systems."
ACM-SIGFIDET Workshop, (1972), pp. 307-322.

Coulouris, Evans, and Mitchell.
"Towards Content-Addressing in Data Bases."
Computer Journal, vol. 15, (May, 1972), pp. 95-98.

Crozzo, D. E. and Kurtz, J. F.
"Building a Base for a Data Base: A Management Perspective."
Datamation, vol. 19, (Oct., 1973), pp. 71-75.

Curtice, Robert M.
"Some Tools for Data Base Development."
Datamation, vol. 20, (July, 1974), pp. 102-106.

Date, C. J.
"Relational Data Base Systems: A Tutorial."
Information Systems (Tou).
New York: Plenum Press, (1974), pp. 37-54.

Dearnley, P.
"A Model of a Self-Organizing Data Management System."
Computer Journal, vol. 17, (Feb., 1974), pp. 13-16.

DeFiore and Berra,
"A Data Management System Utilizing an Associated Memory."
National Computer Conference, (1973), pp. 181-186.

DeFiore, Stillman, and Berra.
"Associative Techniques in the Solution of Data Management Problems."
Proceedings of the ACM, (1971), pp. 28-36.

DeSalvio, Purdy, and Rau.
"Creation and Control of Internal Data Bases Under a FORTRAN Programming Environment."
Communications of the ACM, vol. 13, (Apr., 1970), pp. 211-215.

Dyba, Jerome E.
"Benefits of Data Base and Doing One Better."
Data Management, vol. 11, (Mar., 1973), pp. 31-32.

Eddy, R. and Ramm, M.
"The Role of Data Base Administration."
Proceedings of GUIDE 34, (May, 1972), pp. 224-230.

Engles, Robert W.
"An Analysis of the April 1971 Data Base Task Group Report."
ACM-SIGFIDET Workshop, (Nov., 1971), pp. 69-91.

Erwin, D. B.
"User Experience in Data Base Administration."
Proceedings of GUIDE 34, (May, 1972), pp. 315-337.

Everest, Gordon C.
"The Objectives of Database Management."
Information Systems (Tou).
New York: Plenum Press, (1974), pp. 1-35.

Falor, Ken.
"Survey of Software Packages - Data Base Management Systems."
Modern Data, vol. 4, (May, 1971), pp. 58-59.

Finn, Kenneth W.
"Key Evaluation and Planning Factors for a Data Base Environment - A Case Study."
Data Management, vol. 11, (Sept., 1973), pp. 71-73.

Flynn, Robert L.
"A Brief History of Data Base Management."
Datamation, vol. 20, (Aug., 1974), pp. 71-77.

Fong, Collica, and Marron.
"Six Data Base Management Systems: Feature Analysis and User Experiences."
National Bureau of Standards Technical Note 887, (Nov., 1975), 79 pp.

Frank, R. L. and Yamaguchi K.
"A Model for a Generalized Data Access Method."
National Computer Conference, (1974), pp. 45-52.

Friedman, T. D.
"The Authorization Problem in Shared Files."
IBM Systems Journal, vol. 9, (1970), pp. 258-280.

Frosch, Paul S.
"Experience in Data Base Administration."
Proceedings of GUIDE 35, (Nov., 1972), pp. 215-232.

Fry, Frank, and Hershey.
"A Developmental Model for Data Translation."
ACM-SIGFIDET Workshop, (1972), pp. 77-106.

Fry, Smith, and Taylor.
"An Approach to Stored Data Definition and Translation."
ACM-SIGFIDET Workshop, (1972), pp. 13-56.

Gardner, C. C.
"IMS Tutorial."
Proceedings of GUIDE 34, (May, 1972), pp. 168-223.

Gardner, W. D.
"IMS: New Releases are 'Desirable'."
Datamation, vol. 18, (June, 1972), pp. 79+.

Giannotti, Gene.
"Data Base Integrity."
Data Management, vol. 12, (May, 1974), pp. 22-25.

Grafton, W. P.
"Data Base Recovery with IMS/360."
Data Base, vol. 4, (Spring, 1972), pp. 9-12.

Griffith, Kennedy, MacMillan, and Marini.
"Data Base / Data Communications Review."
Proceedings of GUIDE 34, (May, 1972), pp. 59-167.

Joint GUIDE-SHARE Data Base Requirements Group.
Data Base Management System Requirements.
New York: SHARE, Inc., (Nov.11, 1970), 132 pp.

Harvison, C. W. and Radford, K. J.
"Creating a Common Data Base."
Journal of Systems Management, vol. 23, (June, 1972), pp. 8-12.

Hawryszkiewycz, I. T. and Dennis, J. B.
"An Approach to Proving the Correctness of Data Base Operations."
ACM-SIGFIDET Workshop, (1972), pp. 323-348.

Heindel, L. E. and Roberto, J. T.
"ARPL - A Retrieval Process Language."
Computer Journal, vol. 17, (May, 1974), pp. 113-116.

Hollenbach, Raymond.
"An Application of a Data Base System."
Data Management, vol. 11, (Sept., 1973), pp. 68-70.

Huang, J. C.
"A Note on Information Organization and Storage."
Communications of the ACM, vol. 16, (July, 1973), pp. 406-410.

Huhn, Gerald E.
"The Data Base in a Critical On-Line Business Environment."
Datamation, vol. 20, (Sept., 1974), pp. 52-56.

IBM.
"IBM's Objections to the DBTG Report."
CODASYL Programming Language Committee Proposal, #73003-#73012, (Aug., 1973).

Jardine, D. A.
"A Critical Analysis of Data Base Requirements."
Proceedings of GUIDE 35, (Nov., 1972), pp. 105-120.

Jardine, Donald A. (ed.)
Data Base Management Systems.
New York: American Elsevier, (1974), 279 pp.

Jardine, D. A.
"Data Independence."
Proceedings of SHARE XL, (Mar., 1973), pp. 434-447.

Jervis, B. and Parker, J. L.
"An Approach for a Working Relational Data System."
ACM-SIGFIDET Workshop, (1972), pp. 125-146.

Katzan, Harry Jr.
Computer Data Management and Data Base Technology.
New York: Van Nostrand Reinhold, (1975), 347 pp.

King, Paul F. and Collmeyer, A. J.
"Database Sharing - An Efficient Mechanism for Supporting Concurrent Processes."
National Computer Conference, (1973), pp. 271-275.

Knuth, Donald E.
Sorting and Searching.
The Art of Computer Programming, vol. 3.
Reading, Mass: Addison - Wesley, (1973), 722 pp.

Koch, H. S. and Gotterer, M. H.
"An Independent Data Definition Facility for COBOL and FORTRAN."
ACM-SIGFIDET Workshop, (1972), pp. 57-66.

Kreger, Alan and Nathanson, Janet.
"The Tribulations and Triumphs of GIS."
Datamation, vol. 17, (Oct.15, 1971), pp. 20-25.

Lum, V. Y.
"Multi-Attribute Retrieval with Combined Indexes."
Communications of the ACM, vol. 13, (Nov., 1970), pp. 660-665.

Lum, V. Y. and Senko, M. E.
"Analysis of a Complex Data Management Access Method by Simulation Modeling."
Fall Joint Computer Conference, (1970), pp. 211-222.

Lupien, C.
"GUIDE/SHARE Data Base Management System Requirements," (an overview).
Proceedings of GUIDE 32, (May, 1971).

Lyon, John K.
An Introduction to Data Base Design.
New York: Wiley - Interscience, (1971), 81 pp.

Lyon, John K.
"What is a Database."
FDT Bulletin of ACM-SIGFIDET, vol. 5, (June, 1973), pp. 13-43.

Martin, James.
Principles of Data-Base Management.
New York: Prentice-Hall, (1975), 320 pp.

Martin, George N.
"Data Dictionary/Directory System."
Journal of Systems Management, vol. 24, (Dec., 1973), pp. 12-19.

Mastromano, Frank M.
"Data Base Concept."
Management Accounting, vol. 52, (Oct. 1970), pp. 33-38.

McGee, W. C.
"Some Current Issues in Data Description."
ACM-SIGFIDET Workshop, (1972), pp. 1-12.

Meltzer, H. S.
"Current Concepts in Data Base Design."
Proceedings of SHARE XLI, (Aug., 1973), pp. 451-510.

Minsky, N.
"The Associative - Addressing Capability on Cylinder - Memories and Its Applications - An Abstract."
ACM-SIGFIDET Workshop, (1972), pp. 213-242.

Moreira, DeSouza, Pinheiro, and Fernando.
"Integrating Data Base Management into Operating Systems - An Access Method Approach."
National Computer Conference, (1974), pp. 57-62.

Moulder, Richard.
"An Implementation of a Data Management System on An Associative Processor."
National Computer Conference, (1973), pp. 171-176.

Mullin, James K.
"An Improved Index Sequential Access Method Using Hashed Overflow."
Communications of the ACM, vol. 15, (May, 1972), pp. 301-307.

Mullin, James K.
"Retrieval-Update Speed Tradeoffs Using Combined Indices."
Communications of the ACM, vol. 12, (Dec., 1971), pp. 775-776.

Nerad, Richard A.
"Data Administration as the Nerve Center of a Company's Computer Activity."
Data Management, vol. 11, (Oct., 1973), pp. 26-31.

Nichols, Pat.
"Requirements for a User Language."
Proceedings of SHARE XLI, (Aug., 1973), pp. 514-523

Nievergelt, J.
"Binary Search Trees and File Organization."
ACM-SIGFIDET Workshop, (1972), pp. 165-188.

Nijssen, G. M.
"Common Data Base Languages."
Data Base, vol. 4, (Winter, 1972), pp. 7-11.

Nolan, Richard.
"Computer Data Bases: The Future is Now."
Harvard Business Review, vol. 51, (Sept.-Oct., 1973), pp. 98-114.

Nolan, Richard L.
"Data Base - An Emerging Organizational Function."
National Computer Conference, (1974), pp. 897-901.

Nunamaker, Swenson, and Whinston.
"Specifications for the Development of a Generalized Data Base Planning System."
National Computer Conference, (1973), pp. 259-270.

O'Connell, Michael L.
"A File Organization Method Using Multiple Keys."
Spring Joint Computer Conference, (1971), pp. 534-544.

Olle, T. William.
"MIS: Data Bases."
Datamation, vol. 16, (Nov. 15, 1970), pp. 47-50.

Pan, George S.
"The Characterization of Data Management Systems."
Data Management, vol. 9, (June, 1971), pp. 18-23.

Parker, J. L.
"File Structures and Access Methods for a Geographic Information Retrieval System."
ACM-SIGFIDET Workshop, (1972), pp. 382-394.

Parsons, Dale, and Yurkanan.
"Data Manipulation Language Requirements for Data Base Management Systems."
Computer Journal, vol. 17, (May, 1974), pp. 99-103.

Patterson, Albert C.
"A Data Base Management System."
Proceedings of the ACM, (1971), pp. 197-209.

Patterson, A. C.
"A Data Base Management System."
(Based on work of GUIDE-SHARE DBTG),
Proceedings of SHARE XXXVI, (Mar., 1971), pp. 258-269.

Patterson, A. C.
"Requirements for a Generalized Data Base Management System."
Fall Joint Computer Conference, (1971), pp. 515-522.

Plagman, B. K.
"Data Base Administration - Part I."
Data Processing Manual, (portfolio #03-06-04, 1974).
Philadelphia: Auerbach, 16 pp.

Plagman, B. K.
"Implementation of a Data Base Environment."
Data Processing Manual, (portfolio #03-06-03, 1974).
Philadelphia: Auerbach, 18 pp.

Plagman, B. K.
"An Integrated Corporate Data Base Concept and its Application."
ACM-SIGFIDET Workshop, (1972), pp. 395-419.

Plagman, B. K. and Altshuler, G. P.
"Data Base Management - Concepts for Management."
Data Processing Manual, (portfolio #03-06-01, 1974).
Philadelphia: Auerbach, 21 pp.

Plagman, B. K. and Altshuler, G. P.
"Planning for a Data Base Environment."
Data Processing Manual, (portfolio #03-06-02, 1974).
Philadelphia: Auerbach, 20 pp.

Price, Gerald F. (panel chairman).
"Managerial Considerations of Data Base Systems."
Proceedings of GUIDE 35, (Nov., 1972), pp. 513-521.

Price, Gerald F.
"The Ten Commandments of Data Base."
Data Management, vol. 10, (May, 1972), pp. 14-23.

Raichelson, E.
"A Concept Paper on 'Host' vs. 'Own' Data Manipulation Languages in Military Information Systems."
ACM-SIGFIDET Workshop, (1972), pp. 67-76.

Reich, A.
"GIS/2 Query Facility for IMS Data Bases."
Proceedings of GUIDE 35, (Nov., 1972), pp. 56-104.

Reside, K. D. and Seiter, T. J.
"The Evolution of an Integrated Data Base."
Datamation, vol. 20, (Sept., 1974), pp. 57-60.

Romberg, Bernhard W.
"Data Bases: There Really is a Better Way to Manage Your Files."
Information Systems, vol. 20, (May, 1973), pp. 56-58.

Rosenthal, D. B.
"The Distributed Data Base Concept."
Proceedings of GUIDE 35, (Nov., 1972), pp. 276-288.

Schaefer, Marvin.
"DBL: A Language for Converting Data Bases."
Datamation, vol. 16, (June, 1970), pp. 123-130.

Schneider, Allan H.
"The Data Base Administrator Today."
Proceedings of GUIDE 35, (Nov., 1972), pp. 233-236.

Schubert, Richard F.
"Directions in Data Base Management Technology."
Datamation, vol. 20, (Sept., 1974), pp. 49-51.

Secrest, Richard D.
"The Data Base Administrator."
Proceedings of GUIDE XLI, (Aug., 1973), pp. 524-540.

Senko, Altman, Astrahan and Fehder.
"Data Structures and Accessing in Data Base Systems."
IBM Systems Journal, vol. 12, no. 1, (1973), pp. 30-93.

SHARE Data Base Administration Committee.
SHARE Data Base Administration Committee Report.
SHARE Secretary Distribution #246, (June, 1974), 59 pp.

SHARE Data Base Project - Technical Requirements Committee.
"Technical Requirements Committee Report."
Data Base Project, SHARE, Inc.
SHARE Secretary Distribution #247, (May, 1974).

SHARE Data Base Requirements Group.
Data Base Command Language Project Reports.
Proceedings of SHARE XXXIV, (Mar., 1970), pp. 693-712).

SHARE Information Management Group.
"The Data Base Administrator."
Proceedings of GUIDE XLI, (Aug., 1973), pp. 550-587.

Shemer, J. E. and Collmeyer, A. J.
"Database Sharing: A Study of Interference, Roadblock, and Deadlock."
ACM-SIGFIDET Workshop, (1972), pp. 147-164.

Shneiderman, Ben.
"Bibliography on Data Base Structures."
Data Base, vol. 4, (Winter, 1972).

Shneiderman, Ben.
"Optimum Data Base Reorganization Points."
Communications of the ACM, vol. 16, (June, 1973), pp. 362-365.

Sibley, Edgar H.
"The CODASYL Data Base Approach: A COBOL Example of Design and Use of a Personnel File."
National Bureau of Standards - NBSIR 74-500, (Feb., 1974), 78 pp.

Sibley, Edgar H.
"A Data Definition and Mapping Language."
Communications of the ACM, vol. 16, (Dec., 1973), pp. 750-759.

Smith, D. P.
"A Method for Data Translation Using The Stored Data Definition and Translation Task Group Languages."
ACM-SIGFIDET Workshop, (1972), pp. 107-124.

Stacey, G. M.
"A FORTRAN Interface to the CODASYL Data Base Task Group Specifications."
Computer Journal, vol. 17, (May, 1974), pp. 124-129.

Stevens, W. D.
"The Care and Feeding of Data Base Systems."
SHARE Secretary Distribution #239 & #240.
(Oct. - Nov., 1973).

Stonebraker, M.
"Retrieval Efficiency Using Combined Indices."
ACM-SIGFIDET Workshop, (1972), pp. 243-256.

Tani, Paul Y.
"Comparison of Data Base Management Systems Reports."
Proceedings of SHARE XXXVII, (Aug., 1971), pp. 181-203.

Tellier, H.
"What is a Data Base."
FDT Bulletin of ACM-SIGFIDET, vol. 5, (June, 1973), pp. 13-43.

Tou, Julius T. (ed.)
Information Systems.
New York: Plenum Press, (1974), 500 pp.

Traver, Cheryl M.
"Data Bases - Uncontrollable or Uncontrolled."
National Computer Conference, (1974), pp. 903-908.

Treanor, Richard G.
"Data Management Fact and Fiction."
Data Base, vol. 3, (Spring-Summer, 1971), pp. 17-25.

Uhrbach, H.
"The Role of the Data Administrator."
Proceedings of GUIDE 35, (Nov., 1972), pp. 166-169.

Uhrowczik, P. P.
"Data Dictionary/Directories."
IBM Systems Journal, vol. 12, (1973), pp. 332-350.

Walker, W. A. and Gotlieb, C. C.
"Hybrid Trees: A Data Structure for Lists of Keys."
ACM-SIGFIDET Workshop, (1972), pp. 189-212.

Weinmeister, Carl J.
"The Science of Information Management."
Computers and Automation, vol. 20, (Apr., 1971), pp. 20-24.

Whitney, Kevin M.
"Fourth Generation Data Management Systems."
National Computer Conference, (1973), pp. 239-244.

Wilkes, M. V.
"On Preserving the Integrity of Data Bases."
Computer Journal, vol. 15, (Aug., 1972), pp. 191-203.

APPENDIX C

List of Attendees, Participants and Contributors

Appendix C

List of Attendees, Participants and Contributors

Donald L. Adams Chairman, Auditing	AICPA 1211 Avenue of Americas New York, New York 10036
Charles Bachman Evolving Technology	Honeywell Information Systems 300 Concord Road Billerica, Massachusetts 02154
Richard Baum Evolving Technology	Ohio State University 2024 Neil Avenue Columbus, Ohio 43210
Robert Bemer Chairman, Standardization	Honeywell Information Systems P.O. Box 6000 Phoenix, Arizona 85005
Peter M. Benson Auditing	Coopers & Lybrand Abacus House Gutter Lane London EC2 9AH England
John Berg Proceedings Editor	Institute for Computer Sciences & Technology Systems & Software Division National Bureau of Standards Washington, D.C. 20234
Thomas Bergin Standardization	Veterans Administration (308) 810 Vermont Avenue, N.W. Washington, D.C. 20420
R. E. Blasius Standardization	Director, Standards Division Advisory Bureau for Computing, SSC R1232, Lord Elgin Plaza 66 Slater Street Ottawa, Ontario, Canada K1A 0T5
Susan Brewer Recorder, Evolving Technology	Honeywell Process Computer Division 2255 W. Desert Cove Phoenix, Arizona 85029
Emile B. Broadwin Evolving Technology	Computer Sciences Corporation 650 S. Sepulveda Boulevard El Segundo, California 90245

Milt Bryce Standardization	M. Bryce & Associates P. O. Box 15459 Cincinnati, Ohio 45215
Charles H. Burr Government Regulation	Director Computer Systems Management Division State Budget & Control Board P.O. Box 11488 Columbia, South Carolina 29211
James H. Burrows Government Regulation	Headquarters, U.S. Air Force Directorate of Data Automation Washington, D.C. 20330
Richard Canning General Chairman	President Canning Publications 925 Anza Avenue Vista, California 92083
Robert T. Caravella Goverment Regulation	Federal Trade Commission 6th & Pennsylvania Avenue, N.W. Washington, D.C. 20580
Adolph Cecula, Jr. Auditing	U.S. Civil Service Commission 1900 E Street, N.W. Washington, D.C. 20415
Larry Cohn Contributor, Standardization	IBM Corporation 1501 California Avenue Palo Alto, California 94304
Margaret S. Derby Recorder, User Experience	U.S. Civil Service Commission 1900 E Street, N.W. Washington, D.C. 20415
George Dodd Chairman, Evolving Technology	General Motors Research Labs Computer Technology Department Warren, Michigan 48090
Richard Duckett User Experience	UNIVAC 2121 Wisconsin Avenue, N.W. Washington, D.C. 20007
Thomas Duff User Experience	Honeywell Information Systems P.O. Box 6000 Phoenix, Arizona 85005
Ruth F. Dyke User Experience	U.S. Civil Service Commission 1900 E Street, N.W. Washington, D.C. 20415

Jeffery L. Ehrlich Standardization	Manager, Now Data Base General Electric Credit 260 Long Ridge Road Stamford, Connecticut 06902
Dennis Fife Auditing	Chief Computer Science Section Systems & Software Division Institute for Computer Sciences & Technology National Bureau of Standards Washington, D.C. 20234
Tom Fitzgerald Auditing	Manufacturers Hanover Trust Company 4 New York Plaza New York, New York 10015
Elizabeth Fong User Experience	Systems & Software Division Institute for Computer Sciences & Technology National Bureau of Standards Washington, D.C. 20234
James P. Fry Evolving Technology	Assistant Research Scientist Data Translation Project Graduate School of Business Administration University of Michigan Ann Arbor, Michigan 48109
R. Michael Gall Chairman, User Experience	Associate Director for Information Systems Bureau of Manpower Information Systems U.S. Civil Service Commission Washington, D.C. 20415
Robert C. Goldstein Government Regulation	Assistant Professor of Commerce & Business Administration University of British Columbia Vancouver, B.C., Canada
R. A. Hirschfield Auditing	Computer Audit Systems, Inc. 80 Main Street West Orange, New Jersey 07052
Ted Hollander Auditing	Cullinane Corporation 9360 S. Pleasant Avenue Chicago, Illinois 60620

Joseph A. Iwanski User Experience	Computer Systems & Management Science Mobil Oil Corporation 150 East 42nd Street New York, New York 10017
Seymour Jeffery Host	Chief Systems & Software Division Institute for Computer Sciences & Technology National Bureau of Standards Washington, D.C. 20234
Michael M. Kaplan Evolving Technology	Bell Telephone Laboratories 600 Mountain Avenue Murray Hill, New Jersey 07974
Roger J. Kelly User Experience	Office of Comptroller Municipal Building, Room 622 New York, New York
W. Frank King, III Evolving Technology	Research Division IBM Corporation K55/282 Monterey & Cottle Roads San Jose, California 95193
Albert A. Koch Auditing	Ernst & Ernst 1300 Union Commerce Building Cleveland, Ohio 44115
Richard Kurz Contributor, User Experience	Southern Railway System 125 Spring Street, S.W. Atlanta, Georgia 30303
Robert K. Korfhage User Experience	Department of Computer Science Southern Methodist University Dallas, Texas 75275
Thomas C. Lowe Host	Chief Systems Architecture Section Systems & Software Division Institute for Computer Sciences & Technology National Bureau of Standards Washington, D.C. 20234
Gene Lowenthal Evolving Technology	MRI Systems Corporation P. O. Box 9968 Austin, Texas 78766

Don Lundberg Auditing	IBM Corporation 1501 California Avenue Palo Alto, California 94304
Daniel B. Magraw Keynote Speaker Government Regulation	Assistant Commissioner Department of Administration State of Minnesota St. Paul, Minnesota 55155
Jack Minker Evolving Technology	Chairman Department of Computer Science University of Maryland College Park, Maryland 20742
John Nuxall Auditing	Partner Peat, Marwick, Mitchell & Co. 345 Park Avenue New York, New York 10022
T. William Olle Contributor, Evolving Technology	27 Blackwood Close West Byfleet Surrey, KT14 6PP England
Susan K. Reed Recorder, Government Regulation	Systems & Software Division Institute for Computer Sciences & Technology National Bureau of Standards Washington, D.C. 20234
Jean Sammet Host	President Association for Computing Machinery 1133 Avenue of the Americas New York, New York 10036
Lowell S. Schneider User Experience	Martin-Marietta Mail Code 0422, Box 179 Denver, Colorado
Ben A. Schneiderman Contributor, Evolving Technology	Assistant Professor Computer Science Department Indiana University 101 Lindley Hall Bloomington, Indiana 47401

Michael Senko Evolving Technology	IBM T. J. Watson Research Center P.O. Box 218 Yorktown Heights, New York 10598
Edgar Sibley Evolving Technology	Acting Chairman Department of Information Systems Management University of Maryland College Park, Maryland 20742
Chester Smith Standardization	Pennsylvania State University 214A Computer Building University Park, Pennsylvania 16802
Robert Stone Auditing	AICPA 1211 Avenue of the Americas New York, New York 10036
T. E. Swanson Government Regulation	State Comptroller Data Processing 4th Floor, Lucas Building Des Moines, Iowa 50319
Alan Taylor Recorder, Standardization	Taylor & Torrey 15 State Street Boston, Massachusetts 02109
Charles D. Trigg Chairman, Government Regulation	National Association for State Information Systems P.O. Box 11910 Lexington, Kentucky 40511
Dennis Tsichritzis Evolving Technology	Professor, Computer Science University of Toronto 255 Huron Street Toronto, Ontario, Canada M5S1A7
Luc C. vanZutphen Auditing	Professor, Auditing Free University Amsterdam Van Dien & Company 2 Fizeaustraat P.O. Box 4200 Amsterdam 1006, Netherlands
Victor A. Vella User Experience	Facilities Systems Office NCBC Port Hueneme, California 93043
Ian D. Watson Auditing	Peat, Marwick, Mitchell & Co. 11 Ironmonger Lane London EC2 P 2AR England

Ron Weber Auditing	College of Business Administration University of Minnesota Minneapolis, Minnesota 55455
Harold Weiss Recorder, Auditing	Automation Training Center Suite 17, 11250 Roger Bacon Drive Reston, Virginia 22090
Ewart Willey Standardization	Prudential Assurance Company Ltd. 142 Holborn Bars London EC1N 2NH England
Frederick Withington Contributor, Evolving Technology	Arthur D. Little Corporation 35 Acorn Park Cambridge, Massachusetts 02146

A Contributor is one who could not attend but either submitted a paper or added to the Workshop in a manner that deserves our acknowledgment and thanks.

NBS-114A (REV. 7-73)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBS-SP 451	2. Gov't Accession No.	3. Recipient's Accession No.
4. TITLE AND SUBTITLE DATA BASE DIRECTIONS The Next Steps			5. Publication Date September 1976 6. Performing Organization Code
7. AUTHOR(S) John L. Berg, Editor			8. Performing Organ. Report No.
9. PERFORMING ORGANIZATION NAME AND ADDRESS NATIONAL BUREAU OF STANDARDS DEPARTMENT OF COMMERCE WASHINGTON, D.C. 20234			10. Project/Task/Work Unit No. 640.1119 11. Contract/Grant No.
12. Sponsoring Organization Name and Complete Address *(Street, City, State, ZIP)* National Bureau of Standards Department of Commerce Washington, D.C. 20234			13. Type of Report & Period Covered Final 14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

Library of Congress Catalog Card Number: 76-608219

16. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

What information about data base technology does a manager need to make prudent decisions about using this new technology? To provide this information the National Bureau of Standards and the Association for Computing Machinery established a workshop of approximately 80 experts in five major subject areas. The five subject areas were auditing, evolving technology, government regulations, standards, and user experience. Each area prepared a report contained in these proceedings. The proceedings provide guidance on steps managers should follow to prepare themselves and their organization for the installation of data base management concepts. The auditing working panel noted the increased vulnerability of organizations who integrate their formerly dispersed and redundant files into a data base and suggest actions to address this risk. The technology report noted several promising parallel developments but concluded that the future would see evolving, rather than revolutionary data base progress. Government regulations, particularly the drive for individual privacy rights, were seen to play an important role in determining data base directions and the panel's guidance on cost impact suggest that organizations would experience reduced costs with data base technology. Standards pervaded all issues and were found necessary in several sub-areas of data base technology but the panel saw no immediate likelihood of national data base standards. The user experience working panel noted that data base systems had impacted their organizations to the extent of reconsidering existing data flows, areas of responsibilities, and procedures.

17. KEY WORDS *(six to twelve entries; alphabetical order; capitalize only the first letter of the first key word unless a proper name; separated by semicolons)*

Auditing; cost/benefit analysis; data base; data base management; DBMS; government regulation; management objectives; privacy; security; standards; technology assessment; user experience.

18. AVAILABILITY	19. SECURITY CLASS (THIS REPORT)	21. NO. OF PAGES
[X] Unlimited [ ] For Official Distribution. Do Not Release to NTIS	UNCLASSIFIED	175
[X] Order From Sup. of Doc., U.S. Government Printing Office Washington, D.C. 20402, SD Cat. No. C13.10:451 [ ] Order From National Technical Information Service (NTIS) Springfield, Virginia 22151	20. SECURITY CLASS (THIS PAGE) UNCLASSIFIED	22. Price $2.40

USCOMM-DC 29042-P74

get a line on science and technology. Subscribe to

DIMENSIONS NBS

Whether you're in business, or a teacher, scientist, or consumer, you'll want to keep up with the latest developments in science and technology. DIMENSIONS/NBS, the monthly magazine from the Commerce Department's National Bureau of Standards, can help keep you informed. Every day at NBS, one of the nation's largest research laboratories, scientists seek new answers to a host of national problems, including energy conservation, product safety, metric conversion, and pollution abatement. Their findings, reported each month in DIMENSIONS/NBS, have a direct impact on our daily lives.

Subscription price: $9.45 per year.
Order prepaid from the
Superintendent of Documents,
U.S. Government Printing Office,
Washington, D.C. 20402
SD Catalog No. C13.13

☆U.S. GOVERNMENT PRINTING OFFICE: 1976 210-801/380 1-3

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH reports National Bureau of Standards research and development in physics, mathematics, and chemistry. It is published in two sections, available separately:

- **Physics and Chemistry (Section A)**

Papers of interest primarily to scientists working in these fields. This section covers a broad range of physical and chemical research, with major emphasis on standards of physical measurement, fundamental constants, and properties of matter. Issued six times a year. Annual subscription: Domestic, $17.00; Foreign, $21.25.

- **Mathematical Sciences (Section B)**

Studies and compilations designed mainly for the mathematician and theoretical physicist. Topics in mathematical statistics, theory of experiment design, numerical analysis, theoretical physics and chemistry, logical design and programming of computers and computer systems. Short numerical tables. Issued quarterly. Annual subscription: Domestic, $9.00; Foreign, $11.25.

DIMENSIONS/NBS (formerly Technical News Bulletin)—This monthly magazine is published to inform scientists, engineers, businessmen, industry, teachers, students, and consumers of the latest advances in science and technology, with primary emphasis on the work at NBS. The magazine highlights and reviews such issues as energy research, fire protection, building technology, metric conversion, pollution abatement, health and safety, and consumer product performance. In addition, it reports the results of Bureau programs in measurement standards and techniques, properties of matter and materials, engineering standards and services, instrumentation, and automatic data processing.

Annual subscription: Domestic, $9.45; Foreign, $11.85.

NONPERIODICALS

Monographs—Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

Handbooks—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications—Include proceedings of conferences sponsored by NBS, NBS annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

Applied Mathematics Series—Mathematical tables, manuals, and studies of special interest to physicists, engineers, chemists, biologists, mathematicians, computer programmers, and others engaged in scientific and technical work.

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a world-wide program coordinated by NBS. Program under authority of National Standard Data Act (Public Law 90-396).

NOTE: At present the principal publication outlet for these data is the Journal of Physical and Chemical Reference Data (JPCRD) published quarterly for NBS by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements available from ACS, 1155 Sixteenth St. N. W., Wash. D. C. 20056.

Building Science Series—Disseminates technical information developed at the Bureau on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NBS under the sponsorship of other government agencies.

Voluntary Product Standards—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The purpose of the standards is to establish nationally recognized requirements for products, and to provide all concerned interests with a basis for common understanding of the characteristics of the products. NBS administers this program as a supplement to the activities of the private sector standardizing organizations.

Federal Information Processing Standards Publications (FIPS PUBS)—Publications in this series collectively constitute the Federal Information Processing Standards Register. Register serves as the official source of information in the Federal Government regarding standards issued by NBS pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

Consumer Information Series—Practical information, based on NBS research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace.

NBS Interagency Reports (NBSIR)—A special series of interim or final reports on work performed by NBS for outside sponsors (both government and non-government). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Service (Springfield, Va. 22161) in paper copy or microfiche form.

Order NBS publications (except NBSIR's and Bibliographic Subscription Services) from: Superintendent of Documents, Government Printing Office, Washington, D.C. 20402.

BIBLIOGRAPHIC SUBSCRIPTION SERVICES

The following current-awareness and literature-survey bibliographies are issued periodically by the Bureau:

Cryogenic Data Center Current Awareness Service

A literature survey issued biweekly. Annual subscription: Domestic, $20.00; foreign, $25.00.

Liquefied Natural Gas. A literature survey issued quarterly. Annual subscription: $20.00.

Superconducting Devices and Materials. A literature survey issued quarterly. Annual subscription: $20.00. Send subscription orders and remittances for the preceding bibliographic services to National Bureau of Standards, Cryogenic Data Center (275.02) Boulder, Colorado 80302.

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Washington, D.C. 20234

OFFICIAL BUSINESS

Penalty for Private Use, $300

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF COMMERCE
COM-215

SPECIAL FOURTH-CLASS RATE
BOOK

AMERICAN REVOLUTION BICENTENNIAL 1776-1976

75 YEARS
NBS
1901-1976