DATABASE MANAGEMENT

Objectives, System Functions, and Administration

Gordon C. Everest

DATABASE MANAGEMENT

Objectives, System Functions, and Administration

McGraw-Hill Book Company

New York St. Louis San Francisco Auckland Bogotá Hamburg
Johannesburg London Madrid Mexico Montreal New Delhi
Panama Paris São Paulo Singapore Sydney Tokyo Toronto

CREDITS: Page 7: ANSI definitions, copyright © 1977 Computer and Business Equipment Manufacturers Association (CBEMA): Permission granted November 5, 1985; Pages 7 and 47: Definitions of "data" and "integrity" reprinted by permission of Simon and Schuster, Inc., © 1980 by Simon and Schuster, Inc.; Pages 10 and 11: Quotations from Hanold and DeWan reprinted with permission of *DATAMATION®* magazine, © 1972 by Technical Publishing Company, a Dun & Bradstreet Company, all rights reserved; Page 17: Figure 1-11, reprinted with permission of The Diebold Group, Inc.; Page 178: Figure 5-9, portion of MARK IV® File Definition Form reprinted by permission of Informatics General Corporation; Page 288: Figure 7-7, © 1971 Program Products, Inc., author unknown; Page 386: Figure 11-1, *Computerworld,* September 12, 1977, author unknown; Page 528: Figure 14-9, reprinted with permission of Dr. Larry G. Kersta; Page 559: *Time* magazine excerpts, copyright © 1978, 1982, 1984 Time Inc. All rights reserved. Reprinted by permission from TIME; Pages 625, 642, 643, 654, 681 and 764: Six cartoons by Jim Orton, reprinted with permission; Page 655: Figure 16-7, advertisement reprinted with permission of Aetna; Page 759: Figure 18-9, advertisement as it appeared in 1981, reprinted with permission of Britton-Lee; Page 772: Figure 18-12, reprinted with permission from Pergamon Press Ltd.

This book was set in Times Roman by York Graphic Services, Inc.
The editor was Christina Mediate;
the cover was designed by Joan E. O'Connor;
the production supervisor was Marietta Breitwieser.
Project supervision was done by York Production Services, Inc.
R. R. Donnelley & Sons Company was printer and binder.

DATABASE MANAGEMENT
Objectives, System Functions, and Administration

1 2 3 4 5 6 7 8 9 0 D O C D O C 8 9 8 7 6 5

ISBN 0-07-019781-4

Library of Congress Cataloging in Publication Data

Everest, Gordon C.
Database management.

(McGraw-Hill series in management information systems)
Includes bibliographies and index.
1. Data base management. 2. Management information systems. I. Title. II. Title: Data base management. III. Series.
QA76.9.D3E93 1986 001.64'4 85-11404
ISBN 0-07-019781-4

PREFACE

This book is about managing data in organizations. It presumes the use of computers. Managing data involves the use of a Database Management System (DBMS) on the computer and a human Database Administrator (DBA). Hence the dual thrust of this book into both the technology of DBMS and the administrative dimension of developing, using, and controlling data resources.

Several things are different about this book:

- Takes a *broad, comprehensive* view of DBMS, going far beyond those systems which simply augment programming languages (such as COBOL) with some verbs to manipulate a database one record at a time. In this book, a DBMS includes high-level languages for defining and manipulating databases, producing reports with statistics and graphics, and handling data entry screens and menus. Such a view of DBMS encompasses what have been called "fourth generation languages" and "application development systems."
- Focus on the *functions* of database management rather than the more traditional focus on data structures—hierarchical, network, and relational.
- Strong on the *conceptual*, with a clear organization of material, and understandable explanations of some difficult technical issues in DBMS.
- Strong on the *administrative* aspects of managing data. With this emphasis, some have used the term "Information Resource Management (IRM)" to differentiate it from Database Management.
- Up-to-date with the integration of material as it relates to the use of DBMS on microcomputers, distributed systems, and database machines.

The Importance of Information

Information is the lubricant for the operations and decisions carried out in modern organizations. Through the use of "mind-amplifying" computers we are entering an era of information-intensive production. Harnessing and using information resources can make *both* labor and capital equipment more productive. Information is an increasingly valuable organizational resource that must be managed.

Today, there is substantial disparity between the needs of organizations to manage and make available their data resources and the existing state of the technology in currently available DBMSs. The disparity exists for users at all levels—clerks, managers, analysts, and computer systems personnel who develop and maintain management information systems (MIS).

Purpose and Focus

A basic goal in writing this text is to close the gap between the available DBMS technology and the needs of users. The book should:

stretch DBMS builders,
guide DBMS documenters,
humble DBMS vendors,
armor DBMS buyers,
excite DBMS users,
encourage DBMS administrators, and
educate DBMS students.

This text tells a modern organization how to get a handle on its data resources. The answer lies in appointing a responsible manager of these resources and using computer-based facilities for managing data. This book, then, has two major thrusts: organizational and technological—the management of data resources, and the technology of DBMS.

In 1980, Frost and Sullivan, a market analysis firm in New York, estimated the DBMS market at $137 million, increasing 30% annually to over $1 billion by 1987. Their study calls DBMS "the single most significant software product to come along since the advent of computer operating systems . . . its use in time will permeate the EDP world." A similar study by International Data Corp. (Framingham, MA) in 1983, entitled *The DBMS Marketplace*, predicted that by 1987 the market for DBMS products would climb to $1.75 billion. They further predicted that DBMS programs for desktop microcomputer would capture a 12% share of that market. In another study, Strategic Inc. of San Jose, CA, forecasted a 600% increase in the DBMS market during the 1980s, to over $4 billion by 1989.

For some readers this text will generate great expectations for DBMS. When they seek to acquire a commercially available DBMS, they may become disappointed. The material should help steer an organization away from the poor systems, and give them realistic expectations for the system they do acquire *before they actually use it*. A comparison of any DBMS against the material here will reveal its weaknesses and indicate where the organization will have to compensate for them. At least a company can avoid buying a system because it is popular or oversold. The potential of a well-designed, comprehensive DBMS can make us optimistic about the future. This book seeks to present the full potential of DBMS and the administrative challenges of using a DBMS effectively.

A comprehensive DBMS enables an organization to define and create databases,

to maintain established databases with the timely capture and processing of update transactions, to retrieve data from established databases, to maintain the integrity and quality of the data, and to revise the established databases (and related processes) as needs change. A DBMS enables direct end users to access organizational data resources, provides a building block of functional capabilities used to develop information systems, and provides the tools for an organization to manage and control its data resources.

Audience

This text is aimed at professionals in business and government organizations, and students of business administration and computers preparing to develop or use information systems in organizations. It is for those who have or expect to have a responsibility for or an interest in the effective management of organizational data resources.

For persons who both use and control the use of DBMS, this text is detailed enough to foster appreciation for the more technical aspects of database management. For the technical information specialist it reviews the major objectives, functional capabilities, and organizational issues of database management. The specialist can better understand how technical problems and functional capabilities relate to management needs. The book, then, establishes common understanding between the specialist, and the users and managers of information systems.

Since the mid 1950's, several hundred DBMSs have been developed, with many still commercially available and more still being developed. Even so, all systems fall short of the model presented here. This text aims to stimulate and guide the future development of DBMSs which better serve the organizational need to manage and make available data resources. The industry must strive to perfect a tool for database management which satisfies the fundamental objectives and adequately meets the real needs of modern organizations.

Also since the mid 1950's, several thousand organizations have installed and used DBMSs with varied success. With the advent of microcomputers, this number is being multiplied many times over. This text seeks to encourage and direct the professionals who select and acquire DBMSs, who build application information systems using database management facilities, and who use DBMSs to access organizational data resources.

Premises and Assumed Background

This text is committed to the proposition that it is no longer necessary to know how data is physically stored, manipulated, and accessed before learning how to establish, query, and update databases and to manage data resources. Therefore, the text does not attempt to cover data storage structures and access methods. These methods can differ widely from one system to another and learning them can quickly bog the student down in unnecessary detail. Most users do not (at least should not) need to know how data is stored on secondary devices and accessed within the computer system. Instructors who feel their students do not adequately appreciate basic storage structure concepts may

add outside readings—including a look at the storage structure for the DBMS available to the students for class projects. Understanding how COBOL stores and accesses data files can provide a useful example but would not be necessary. It would enable those who already know COBOL to relate that knowledge to the material in the text.

This text also takes the position that it is no longer necessary to know high-level programming languages (such as COBOL and FORTRAN) before learning about databases, and high-level *data* definition languages and *data* processing languages. This extends the accepted view that you need not learn assembly language programming before learning a high-level programming language. An average high-level language programmer using an optimizing compiler can usually produce better object code at less cost than can an assembly language programmer. Comprehensive, high-level data languages for defining and manipulating data, and effective DBMS, eliminate the need to learn low-level *data* languages first. Only the specialist concerned with the efficiency of a specialized process beyond the capabilities of the DBMS needs low-level languages. Unfortunately, many of today's DBMSs lack comprehensive functional capabilities and high-level data languages.

This text has been used experimentally in colleges over ten years. Students are assumed to have a background in computers and data processing (as found in an introductory text, e.g., Davis, *Computer Data Processing*, McGraw-Hill, 1973), with some appreciation for the role of information systems in organizations (as in Davis and Olson, *Management Information Systems*, McGraw-Hill, 1985).

This text seldom mentions specific DBMSs. Its focus on general concepts, functional capabilities, and languages transcends the ever changing software marketplace. Readers already familiar with a DBMS can easily relate this text to their own understanding and experience. To enhance learning, an instructor can supplement this text with the manual for an existing DBMS. The student can see a consistent, practical example of a DBMS and can use the system to design, create, query, and manipulate databases in class projects. The Instructor's Manual contains suggestions for using this text and constructing a course on data management.

Contrast with Existing DBMS Literature

To those somewhat familiar with the literature, this text will initially appear to be out of the mainstream of thought in database management technology. For example, the text is not organized around what many consider to be the major basis for classifying DBMS: hierarchical, network, and relational data structures. Readers are cautioned not to reject the approach taken here without examining the arguments and the alternative classification presented. The chosen data structure classes must be rooted in fundamental not spurious differences, and must be related to their corresponding, high-level data languages. The taxonomies developed in this text were carefully thought through. Most students of DBMS will readily accept the classifications used herein and find them superior to existing ones.

Consider how this book differs from current database management texts:

- Emphasis on concepts and principles.
- Focus on the logical aspects of data structures, with minimum discussion of physical data storage and access methods.
- Description of logical data structures which transcends the minor differences between "network" and "relational" data structures.
- Focus on high-level data languages for definition and manipulation, with little attention to low-level, one-record-at-a-time navigational languages.
- Little discussion of existing DBMSs which become obsolete quickly, or represent at best a since-improved-upon technology.
- Emphasis on the database management functions to serve the needs of end users and management.
- Broad coverage of organizational acquisition, use, and administration of DBMS.

Organization of the Chapters

After an introductory chapter, this text is organized into three parts. Chapter 1 answers the "why" and "so what can I hope to achieve" with database management technology.

Part I provides both a foundation and overview of DBMS by exploring several different definitions and classifications of DBMS.

Chapter 2 identifies several factors which motivate an organization to consider DBMS. It outlines the objectives an organization would hope to achieve by applying database technology, perhaps acquiring a DBMS, and embracing the philosophy and principles of database management. It concludes with a brief look at the evolution of the database approach to managing data in organizations.

Chapter 3 develops a conceptual model of DBMS based upon different user types, and the functions performed. This chapter also relates DBMS to other types of systems oriented to processing other forms of data, and relates a DBMS to other functions within a computer environment.

Chapter 4 differentiates DBMS based upon the underlying data structure class (or "data model") which can be defined with a DBMS.

Chapter 5 explores differences in DBMS based upon user interface, language, and modes of operation.

The casual reader or one new to the field can focus on the development of the conceptual model in Chapter 3, skimming the rest of the material in Part I. Those previously exposed to database literature or familiar with DBMS will want to read Part I in more detail to set their present state of knowledge in perspective.

Part II consists of nine chapters covering the major technological functions of a comprehensive DBMS.

Chapter 6 focuses on the design and definition of a database structure.

Chapters 7 and 8 describe database retrieval using high-level facilities, first on a single flat file, then extended in Chapter 8 to retrieve from hierarchical and multifile structures.

Chapter 9 covers the functions for getting data into the database—initial creation and ongoing update—still using high-level functional capabilities.

Chapter 10 considers the functions and interface for programming users who build an information system using the DBMS as a development tool.

Chapter 11 discusses the meaning and means of data independence, an important characteristic of DBMS providing for evolvability. Data independence implies a difference in the way data is viewed, and therefore, the need for a family of data conversion processes. Once the database has been established and used for a time, the revision process becomes increasingly important.

Chapters 12, 13, and 14 cover the DBMS functions for maintaining database integrity—backup and recovery, quality control through data validation and concurrent update control, and access control through identification, authorization, encryption, and threat monitoring.

Part III looks at the administrative side of database management.

Chapter 15 details the organization, functions, and tools of database administration.

Chapter 16 explores the legal issues pertaining to database administration, particularly the organizational response to laws relating to the right to privacy, due process, and public access.

Chapter 17 gives some guidelines for evaluating, selecting, and acquiring a DBMS.

Finally, Chapter 18 examines future developments in database technology, including distributed databases, database machines, and standards activities.

Each chapter ends with a summary, a set of exercises for students to test their own comprehension and to apply and extend the material in the chapter, and a selected, annotated bibliography for further reading. The text extensively uses figures to organize, exemplify, and summarize. Each figure with its caption should be self-explanatory. The figures and accompanying captions provide a unique and efficient way of overviewing or reviewing the major concepts in the text.

A special thanks to those who took the time to review this manuscript or portions thereof in various stages of its development: Gordon B. Davis, University of Minnesota; Donald L. Davis; James C. Emery, University of Pennsylvania, for his early encouragement; David Jefferson, NBS; Hugh F. Juergens, formerly University of Wisconsin, Madison; Salvatore March, University of Minnesota; G. M. Nijssen, University of Queensland, Australia; T. William Olle; Craig Roger; George M. Scott, University of Connecticut, Storrs; James A. Senn, Georgia State University; Ben Shneiderman, University of Maryland; James L. Smith, Northwestern College; Ronald Teichman, Pennsylvania State University; Yannis Vassiliou, New York University; Eric A. Weiss; and Gio Wiederhold, Stanford University. I am also indebted to many former graduate students at the University of Minnesota who provided detailed feedback on prior versions of this text, as well as attendees at the University of Michigan Summer Engineering Conferences. Finally, a special thanks to Nancy Rudisill and Mary Jo Gregory at York Production Services for their patience and their responsiveness to my requests.

The pioneering emphases of this text in a relatively new field make feedback very important. You are encouraged to write or call with any comments or suggestions.

Gordon C. Everest

CONTENTS

Preface xi

1 Managing Data in Organizations 1

Four Components of a Data Processing System 3
Data and the Database Approach 7
Managing Data Resources 13
Organizational Context for Database Management 16
The State of Database Management in Organizations 18

Part I Overview of Database Management Systems

2 Motivation, Objectives, and Evolution of the Database Approach 27

Traditional Approach to Application System Development 28
Motivations toward the Database Approach 31
Countervailing Forces Inhibiting DBMSs 34
Objectives of Database Management 36
Evolution of Database Management Systems 52

3 A Conceptual DBMS Model 73

Several Dimensions of DBMS 74
User Roles in the DBMS Environment 74
Database Management Functions 84
A Conceptual DBMS Model 88
How the DBMS Model Satisifes 97
The DBMS Environment 106

4 Logical Data Structures 119
A Taxonomy of Data Structures 120
Single Flat File Data Structure 123
Single Hierarchical Data Structure 129
Multifile Data Structure 133
Object-Relation Data Structure 143

5 User Interface, Language, and DBMS Operation 159
Evolution of the User Interface 160
Principles of Direct Manipulation 163
Types of Online Users: By Usage Mode and Dialogue Style 168
Additional Online Support 181
Types of Data Languages: Definition, Manipulation, Mapping 186
Data Languages on Data Structures 187
DBMS Modes of Operation 188

Part II Database Management System Functions

6 Database Design and Definition 197
The Process of Database Design 198
The Process of Database Definition 205
Database Definition Information 213
Importance of Formal Database Definition 242
Logical Database Design 244

7 Database Retrieval: Flat Files 257
Steps in the Retrieval Process 258
Selection 265
Projection and Derivation of Data Items 273
Ordering, Control Breaks, and Statistics 275
Formation and Presentation of Output Results 280
A General Flat File Retrieval Language 287
Graphical Presentation of Statistical Output 294
Natural Language Query Systems and Artificial Intelligence 299

8 Database Retrieval: Hierarchical and Multifile Structures 307
Development of Retrieval Languages on Hierarchical Data Structures 308
Querying a Single-Path Hierarchical Data Structure 311
Querying a Multipath Hierarchical Data Structure 323
Retrieval from a Multifile Data Structure 330

9 Database Creation and Update 339
Data Capture and Mechanization 340
Database Creation 341
Transaction Processing and Database Update 346
General Update Operations 354

10 Programming User Facilities for System Development 359
Methods of Program Access to a Database 361
COBOL Input-Output Statements 362
Program Communication with the DBMS 364
Form, Time, and Scope of Program-System Communication 371
Additional Facilities for the DBMS System Programmer 375
Converging Facilities for Programming and Nonprogramming Users 376

11 Data Independence, Data Conversion, and Database Revision 385
Historical Evolution of the Program-Data Relationship 387
Data Independence 390
Binding, Data Independence, and Evolvability 398
Data Conversion Processes 409
Database Revision 416

12 Database Integrity: Backup and Recovery 427
Overview of Integrity Control Functions 428
The Processes of Database Backup and Recovery 430
Backup Strategies 431
Summary Chart of Traditional Backup and Recovery Strategies 439
Residual Dump Backup Strategy 441
Variables in the Backup Process 444
Process Checkpoint and Restart 446

13 Database Integrity: Quality Control and Concurrent Update 453
Data Validation 454
Update Authorization 465
Concurrent Update Control 466
Update Synchronization 496

14 Database Integrity: Access Control and Encryption 505
Data Access Control Policies and Approaches 508
A General Model of Data Access Control 510
User Identification and Authentication 515
Authorization 529
Controlling Inferences from Statistical Data 538
Encryption 544
Threat Monitoring and Audit Trail 562

Part III Database Administration

15 Database Administration Organization, Functions, and Tools 575

History of and Need for Database Administration 576
Database Administration within the Organization 576
Organization of the Database Administration Function 586
Functions of Database Administration 589
Database Administrator Tool: The Data Dictionary 601
DBMS Performance Monitoring and Usage Statistics 610
Organizational Response to a Database Administration Role 613

16 Data Privacy and Fair Information Practices 623

Privacy and Personal Data 624
The Three Basic Rights 628
Organizations, Individuals, and Personal Data 633
Legislative Approaches to Fair Information Practices 640
Organizational Response 653

17 DBMS Selection and Acquisition 671

Beginning the Search and Evaluation Process 672
Acquisition Alternatives 679
DBMS Evaluation Process 685
Investigating DBMS Packages ("Getting Answers") 693
Selection Criteria ("Asking Questions") 697
Final Selection, Contract Negotiation, and Acquisition 704
Installation and Use 714

18 Trends to the Future in Database Management 731

DBMS Development and Usage Trends 732
Organizational Resistance to New DBMS Tools 736
Databases in a Distributed Processing Environment 738
Backend Database Management Machine 750
Emerging Standards in Database Technology 760

Index 783

CHAPTER

ONE

MANAGING DATA IN ORGANIZATIONS

1.1	FOUR COMPONENTS OF A DATA PROCESSING SYSTEM	3
1.1.1	The Problem of Data	3
1.1.2	A Copernican Revolution in Data Processing	4
1.2	DATA AND THE DATABASE APPROACH	7
1.2.1	Data versus Information	9
1.2.2	Data versus "Library" Information	11
1.2.3	The Singularity of "Data"	11
1.2.4	Database	11
1.2.5	The Database Approach	12
1.3	MANAGING DATA RESOURCES	13
1.3.1	The Human Component: Database Administrator (DBA)	14
1.3.2	The Machine Component: Database Management System (DBMS)	14
1.3.3	Overview of Relationships in Database Management	15
1.4	ORGANIZATIONAL CONTEXT FOR DATABASE MANAGEMENT	16
1.4.1	Managers and Organizations	16
1.4.2	Management Information Systems	17
1.4.3	Role of the Database in an Organization	18
1.5	THE STATE OF DATABASE MANAGEMENT IN ORGANIZATIONS	18
	SUMMARY	20
	EXERCISES	21
	SELECTED REFERENCES	22
	Management Information Systems—the DBMS Context	22
	Books and Articles Generally on DBMS	22
	Serials Devoted Substantially to DBMS	26

"I know you have last year's employment statistics somewhere in the computer system. Why can't you get them for me quickly?"

"Why must I always consult a programmer to get the data I want?"

"I just want to know how many electrical engineers are in the company, where they are in the organization, and the projects they're currently working on. Why do I have to specify all this formatting information and tell the system how to search for the data?"

"It would sure be nice if I could write my own little computer program for my unique data needs; the existing facilities are so awkward and inefficient."

"Don't expect me to use the data from the corporate database. It's old and inaccurate; I can't depend on it to do my job properly."

"Why can't I simply tell the system what the data is supposed to look like and have the system ensure that the actual data conforms?"

"With all that effort just to revise the database structure and add a few more items, I'll build my own database for this new application, even though it's redundant."

Complaints of this sort are common in most organizations today and reflect a need to manage and make available data resources. A well designed, comprehensive database management system (DBMS) can fulfill these needs. The principles of database management are universal, even if an organization chooses not to use a DBMS.

Data is an important resource in the operation and management of an organization. Effective data resource management makes an organization responsive and successful. Through a base of data, the organization has an *image* of itself and its environment, its past and its future. This image stimulates and conditions the response of the organization. With a timely and accurate image, daily operations run more smoothly. People in the organization make better decisions and respond more confidently to organizational change and environmental demands.

This text looks at data within an organizational context, including both technical and administrative considerations—the machine aspects and the human aspects of data management.

Gathering and using data in organizations is not new. However, the recent use of computers has forced greater discipline on the management of data. The computer enables an organization to better respond to internal demands for data as well as external demands from governmental and public agencies. After nearly three decades, the computer industry and users have gained solid experience in using computers to process and manage data. This text synthesizes this past experience and indicates future directions for development.

This chapter introduces major database concepts and terminology. Later chapters build on this base and context for understanding. Remaining sections in this chapter discuss the meaning of data, database, the database approach, database management, database management systems, database administration, and finally, the context of DBMS—organizations, managers, and MIS.

1.1 FOUR COMPONENTS OF A DATA PROCESSING SYSTEM

A computer-based data processing system has four basic components (see Figure 1-1): machines, programs, data, and people. Machines and programs are also called "hardware" and "software."

The invention of modern computers during the late 1940s led to an industry that produced and marketed such machines.

As business organizations responded to the computer revolution, they concentrated on selecting and installing equipment, training programmers and analysts, and mechanizing existing office procedures. They devised sophisticated algorithms to extend the numerical and scientific capabilities of the new machines, developed higher-level languages to more efficiently utilize human resources, and evolved supervisory and operating systems to more efficiently use the machine. Both the computer industry and using organizations have worked hard to develop better programming techniques (to tell a computer what to do). High-level programming languages have been developed to such a degree that today the definition of a process* can be quite machine independent.

1.1.1 The Problem of Data

Data are "facts" represented by values—numbers, character strings, or symbols which carry meaning in a certain context. These values can be punched into cards, stored on secondary storage devices (such as magnetic tape or disk), or stored in the central memory of a computer. Programs direct the movement and manipulation of data within a computer system.

Since the mid-1950s computer professionals have striven to develop programming languages independent of machines and geared to users. Yet they made relatively little effort to do the same for data. Only recently has there been a significant effort toward developing machine-independent, high-level data languages. This effort has been partly fostered by the emergence of the microcomputer industry. Could the computer industry have focused on data first? Probably not. Effective computer use required an initial emphasis on programming and programming languages because without programs the computer does nothing.

Figure 1-2 illustrates the relationships between machines, programs, data, and people. Notice that data is not well decoupled from programs which use it or machines which store it. This is the problem of data in most organizations today.

Several problems stem from data dependent programs and machine dependent

*"Process" is used in this text in a very precise sense: a sequence of operations scheduled and executed over time. A computer *program* (or set of programs) defines a process to be carried out on a machine. A program consists of a sequence of statements or commands written in some programming language. A *procedure* defines a process to be carried out by a person. An intuitive definition of *process* is provided by Butler W. Lampson, "A Scheduling Philosophy for Multiprocessing Systems," *Communications of the ACM* (11:5), 1968 May, page 347; and an extended, formal definition is given by J. J. Horning and B. Randell, "Process Structuring," *Computing Surveys* (5:1), 1973 March, pages 5-30.

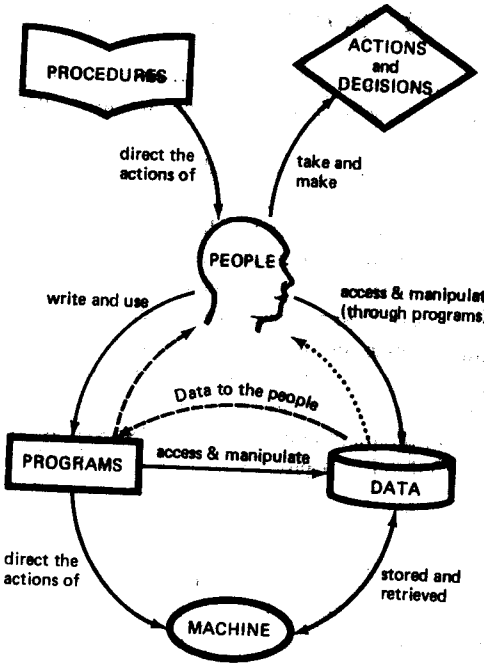

Figure 1-1. Components of a Data Processing System.

Until recently, the focus has been largely on the machine and the programs it executes, perhaps explaining the popularity of the terms "hardware" and "software."

data. Sharing data among applications and transferring data across machines and systems create data problems. When new systems are installed to replace existing hardware and/or software, organizations must transfer (convert) their existing programs and data over to the new system. People in the computer industry are predicting data conversion problems more extensive than program conversions of the past. Data is rapidly being converted to machine-processable forms. Moving to a new computer system now involves converting both programs (made easier with high-level languages) and large volumes of data. As our investment in machinable data increases, the problem of data increases the cost of shifting to new machines and to new technology, and of incorporating new application systems.

The technological development of data processing systems should eventually reconcile programs and data to serve people and meet their needs. At the same time programs and data must become mutually independent as well as machine independent (Figure 1-3). The growing interest in data, data structures, database design, high-level data languages, menu-driven facilities for the online database user, and managing data resources in organizations suggests such a trend.

1.1.2 A Copernican Revolution in Data Processing

The problem of data is causing a Copernican revolution in data processing. At first the profession seemed to focus exclusively on programs, viewing input and output of