

**JAMES MARTIN**

**PRINCIPLES  
OF DATA-BASE  
MANAGEMENT**

# OF DATA-BASE MANAGEMENT

**JAMES MARTIN**

*IBM Systems Research Institute*

**PRENTICE-HALL, INC.,** Englewood Cliffs, New Jersey



*Library of Congress Cataloging in Publication Data*

MARTIN, JAMES

Principles of data-base management.

Includes bibliographical references and index.

1. Data base management. I. Title.

QA76.9.D3M37      001.6'442      75-29054

ISBN 0-13-708917-1

**Principles of Data-Base Management**

*James Martin*

© 1976 by Prentice-Hall, Inc.  
Englewood Cliffs, N.J.

All rights reserved. No part of this book may be reproduced in any form, by mimeograph or any other means, without permission in writing from the publisher.

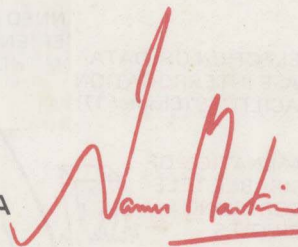
10 9 8 7

Printed in the United States of America

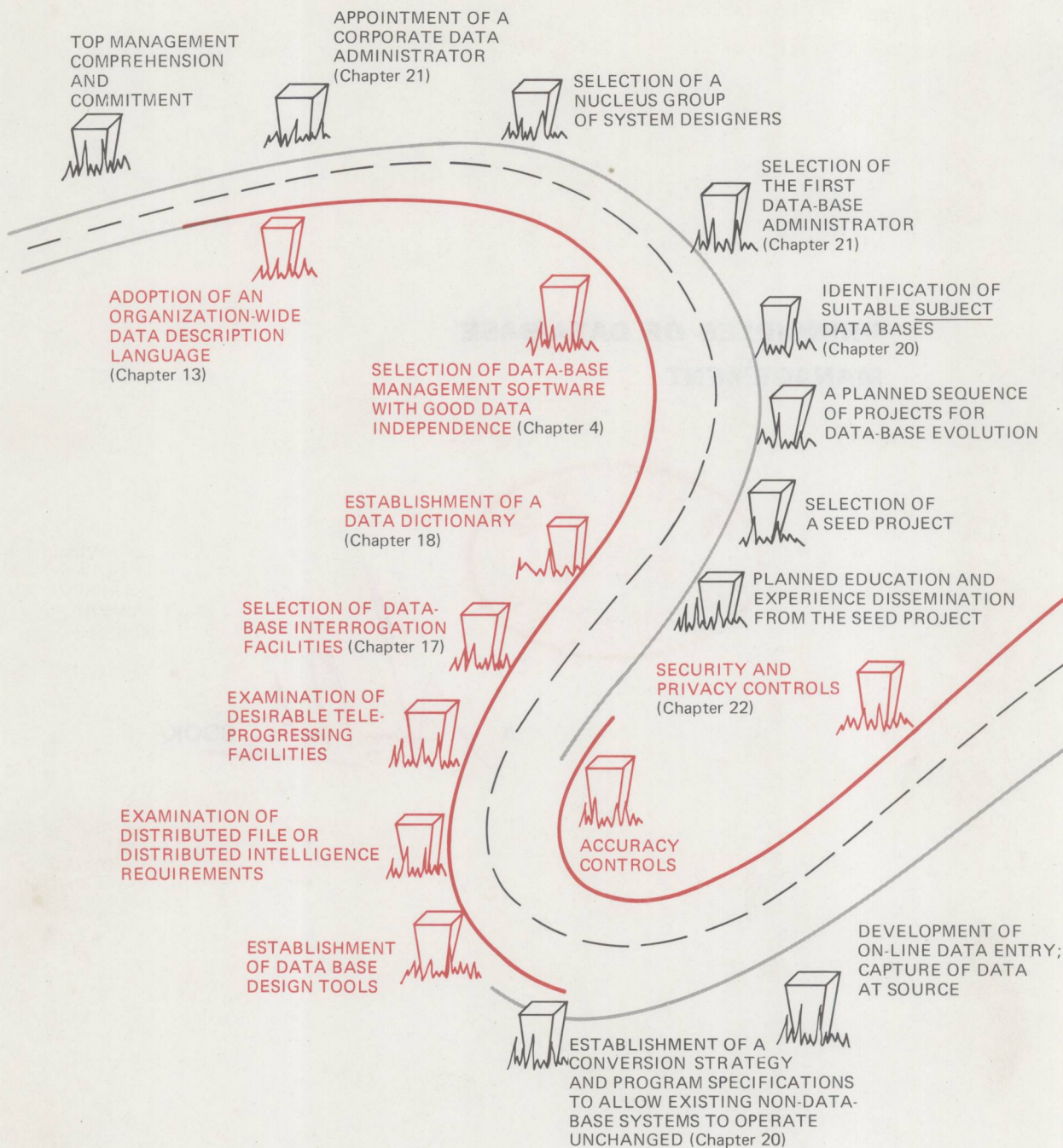
PRENTICE-HALL INTERNATIONAL, INC., *London*  
PRENTICE-HALL OF AUSTRALIA, PTY. LTD., *Sydney*  
PRENTICE-HALL OF CANADA, LTD., *Toronto*  
PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*  
PRENTICE-HALL OF JAPAN, INC., *Tokyo*  
PRENTICE-HALL OF SOUTHEAST ASIA (PTE.) LTD., *Singapore*

# **PRINCIPLES OF DATA-BASE MANAGEMENT**

A

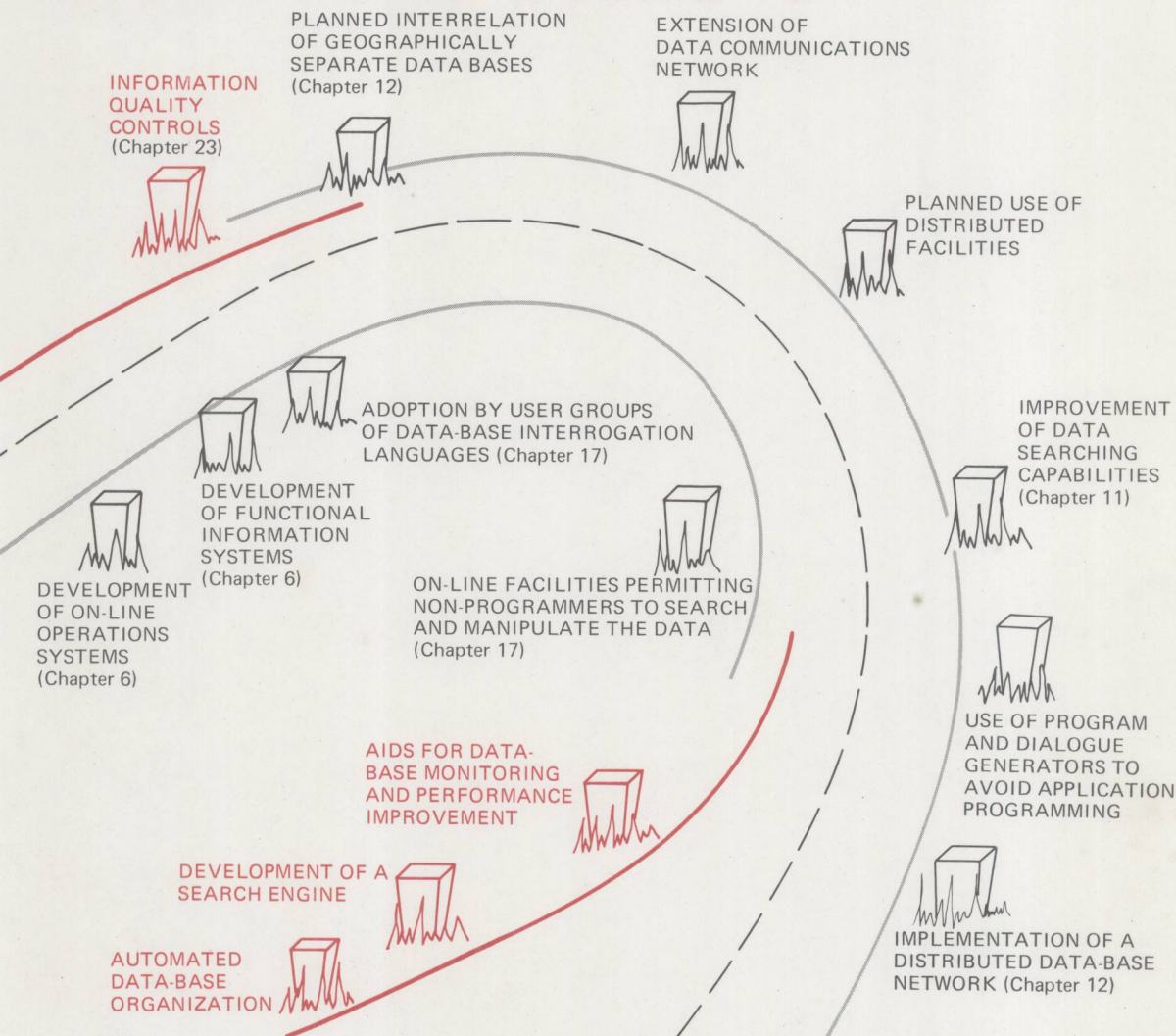
A red ink signature that reads "Vamsi Martin". The signature is stylized, with a large, sweeping 'V' and a cursive 'Martin'.

**BOOK**



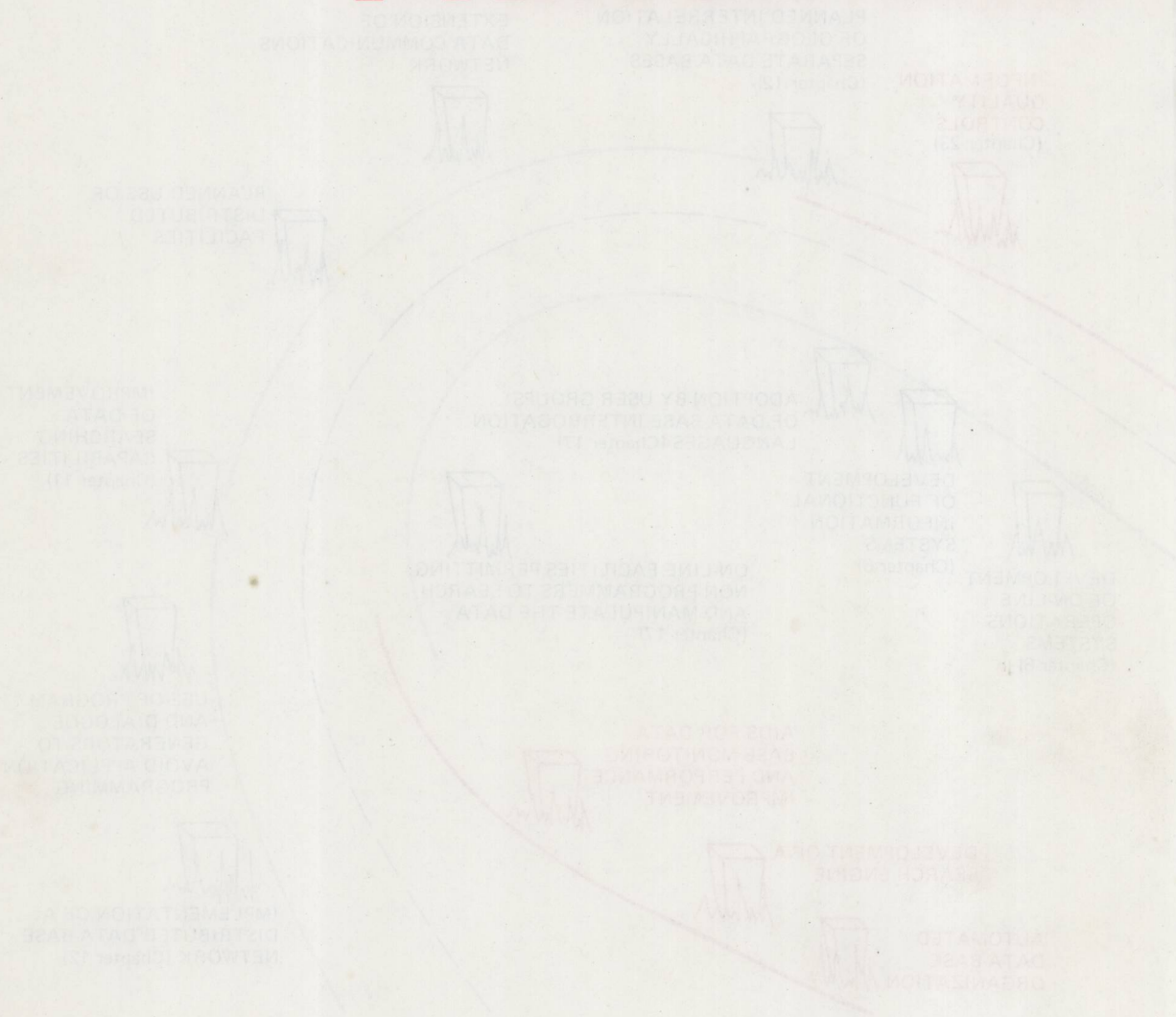


# MILESTONES ON THE ROAD TO A COMPREHENSIVE INFORMATION SYSTEM



WILKINSON ON THE ROAD TO A  
COMPREHENSIVE INFORMATION SYSTEM

# PRINCIPLES



**TO CHARITY**



## **PREFACE**

### **INTENT**

One of the most badly needed courses in universities, business schools, and other establishments which teach computing is a course on the realities of data-base technology. The 1970's is the decade of the data base. Probably the biggest difference between the next generation of computers and the present will be massive on-line storage and its software. By the end of the 1970's, much of the computing in industry and government will relate to the data bases which have been painfully constructed piece by piece, and management effectiveness will relate to the quality of their organization's data sources and the versatility with which they can be used.

At the time of writing, data base technology is widely misunderstood. Its role as the foundation stone of future data processing is often not appreciated. The techniques used in many organizations contain the seeds of immense future difficulties. Data independence is often thrown to the winds. Data organizations in use prevent the data being employed as it should be. And most educational establishments do not yet teach a data base course. This book is being used at the IBM Systems Research Institute as the text for such a course.

The book covers the topics of logical and physical data structures, data security, accuracy and privacy, and dialogues with a data-base system terminal, in what is considered sufficient detail for an introductory text. These topics are covered in more detail in some of the author's books listed on the front endpapers.

### **ACKNOWLEDGEMENTS**

The author wishes to thank many students who have reviewed the text critically. He is very grateful for the detailed comments from Mr. R. M. Gale, Dr. E. F. Codd, Mr. R. W. Holliday, Mr. H. S. Meltzer and Mr. John Mahony of AT&T. The author is especially grateful to his wife for her assistance, and to Miss Cora Tangney who helped with the manuscript. Last, and most important, he would like to thank the late Dr. Kopley whose encouragement will bear fruit for decades.

JAMES MARTIN

## CONTENTS

Preface *xiii*

Index of Basic Concepts *xv*

### PART

## WHY DATA BASE?

- 1** Corporate Data Bases 2
- 2** Data Basics 8
- 3** Categories of Data Usage 23
- 4** Flexibility and Independence 35
- 5** The Changing View of Data 48
- 6** Operations Systems versus Information Systems 65

PART



**DATA ORGANIZATION**

- 7** Schemas and Subschemas 74
- 8** Tree and Plex Structures 85
- 9** Relational Data Bases 95
- 10** File Addressing 111
- 11** Searching 122
- 12** Distributed Data Bases 136

PART



**DATA-BASE SOFTWARE**

- 13** Types of Data-Base Languages 148
- 14** Data-Base Management Systems 166
- 15** The CODASYL Data Description Language 175
- 16** IBM's Data Language/I 191
- 17** Query Languages 205
- 18** Data Dictionaries 221



- 19**      Infrastructure    230
- 20**      Evolution    235
- 21**      The Data Administrator and Data-Base Administrators    260
- 22**      Security and Privacy    271
- 23**      Information Quality    292
- 24**      Management Information Systems    300
- 25**      Resistance to Rational Systems    310
- 26**      Epilogue: How to Succeed    319
- 
- Glossary    326
- 
- References    341
- 
- Index    345

## INDEX OF BASIC CONCEPTS

*The basic concepts, principles, and terms that are explained in this book are listed here along with the page on which an introductory explanation or definition of them is given. There is a complete index at the end of the book.*

Addressing	111	Data description language	162
Area (CODASYL term)	188-9	Data dictionary	159, 221-2
Associative memory	121	Data element	5
Attribute	19-20	Data independence	45
Authorization schemes	284-5	Data item	5
		Data manipulation	
Binary search	112-3	language	155-7, 172
Block search	111-2	Data migration	62
Bucket capacity		DDL (CODASYL data	
(with hash addressing)	119-21	description language)	175
Chained organization	128-30	Device media control language	153
Check points	277	Dictionary (data)	221-2
CODASYL	148, 176	Direct-access storage	9-15
Complex mapping	82	Distributed data base	136
Complex plex structure	88, 90	Distributed intelligence	138
Corporate standards	165, 175	DL/I (IBM Data Language/I)	191
Cycles	90-1	Domain	97-8
		Elementary item	5
Data administrator (corporate)	260	Entity	19
Data aggregate	177-8	Entity identifier	29
Data base	4	Executive vs. supervisory system	66
Data-base administrator	260		
Data-base interrogation		Father and grandfather files	54
language	157-8	Field	5
Data-base management system	166-7	File	19
Data Base Task Group (DBTG)	148	Flat file	98
		Forms-oriented language	207-10

Global logical view (of data)	56	Operations system	67-9
Grandfather and father files	54	Password	288-9
Hashing (addressing method)	119-21	Physical data independence	56
Heuristic searching	206	Physical data structure	21
Hierarchical file	87	Physical DBD (data-base description)	191
Index	114	Plex structure	87-8
Indexed nonsequential file	114-7	Primary key	28
Indexed sequential file	114	Privacy	271
Information system	67-9	Privacy lock	283-4
In-line processing	22	Program specification	
Instance of record (etc.)	75	block	204, Fig. 16.11
Interrogation language	205	Projection (in a relational data base)	106
Intersection data	201	Query language	205
Inverted file	29	Random-access storage	8-9
Inverted list	29	Randomizing	
Journals	279-80	(addressing method)	119-21
Key	28	Real-time processing	27
Key (for privacy)	287-9	Record	5
Key conversion (addressing)	117-9	Relation	96
Levels of management information	300-2	Relational data base	96
Lock (privacy)	283-4	Schema	74-5
Logical child (in DL/I)	193-5	Schema description language	153-4
Logical data-base record (in DL/I)	193	Scrambling (addressing method)	119-21
Logical data independence	56	Searching (heuristic)	206
Logical data structure	21	Secondary index	128-9
Logical DBD (data-base description)	196	Secondary key	28-9
Logs	279-80	Security	271
Loops	90-1	Seek	8
Mapping (simple and complex)	82	Segment	192
Migration of data	62	Sensitivity (of program to data)	204
Multilist organization	132-3	Sequential-access storage	8-9
Multiple-key retrieval	122-8	Set (CODASYL term)	178, 180
Network	87	Subschema	76, 80
Normalization	96	Subschema description language	154-7
Off-line	8	Supervisory vs. executive system	66
On-line	8	Symbolic address	117
		Tree	85-6
		Tuple	96



## WHY DATA BASE?

Knowledge is power.

Francis Bacon

He who runs the information runs the show.

*Joseph Goebbels*

# 1

## CORPORATE DATA BASES

It is pleasant, now that the world of computers has become so complex, to let one's imagination drift back to the days of Dickens. In those days data processing was done by a clerk with a quill pen who was perched on a high stool and perhaps wearing a top hat. In front of him he had a set of thick and well-bound ledgers. If an order were made for a certain quantity of goods, a clerk would deal with this transaction in its entirety. He might look at his stock sheets to see whether the order could be fulfilled from stock or whether some of it had to be manufactured. He would update the order book, and if any goods were sent, he would modify the stock sheets to make out a bill for the customer and make an entry on the appropriate page of a customer ledger—a simple process which was easy to understand. If anyone had any query about the state of the business, about a certain item of stock, or about an outstanding debt of a customer, the clerk could turn to the appropriate pages of his ledgers and immediately produce the answer. One can imagine such a clerk today taking orders by telephone and answering queries over the telephone. He would balance his books at the end of each day, and if costing figures were required, he could maintain them so that they were as up-to-date as required.

However, admirable as the methods of the Dickensian clerk were, they could work only in a fairly small company. As the company grew, the size of the ledgers increased until several clerks were needed to maintain them. Division of labor made the job easier, and one clerk would maintain the stock sheets while another did the billing, and so forth. Earlier in this century various means for mechanization were introduced, and to make efficient use of them, the work was split up into batches. For example, several hundred transactions may have been grouped into a batch. One accounting function would be carried out on all the transactions by one clerk or one machine, and then the next function would be performed by another clerk or another machine. When punched-card accounting was introduced



it became economical to have very large batches. Many trays of cards would be fed through one machine before the setup of that machine was changed for the next function it would perform. Similarly, with the use of magnetic tape on computers, large tape files would be processed with one program before the file was sorted and made ready for the next operation. In working this way the flexibility of the old clerical methods was lost. It was no longer possible to give one transaction individual treatment. It was no longer possible to give quick answers to inquiries about status of the account, or the credit worthiness of a customer, or the amount of an item in stock. Or, at least, if such an inquiry was made, the answer might be a week or more out-of-date. When items were to be posted it was necessary for the computer to read every item in the file as it scanned its way to the ones to be updated, and every item had often to be written out afresh on a new tape, whether it was updated or not.

Batch processing with data rigidly divided into separate files for each application was not the ideal way to operate. It would have been much more convenient for management to have all the information about running their organization up-to-date and at their fingertips. Because of the nature of data-processing techniques, management was living with a compromise. Today the compromise has been in existence for so long that it has become the accepted method of operation, and little thought is given to its desirability.

The use of a data base is like having a superbly fast and brilliant Dickensian clerk who keeps data for many applications. He organizes his books so that minimum writing is necessary and so that he can search the books quickly to answer any queries that may come along. Unlike his pedestrian predecessors who could write or read items in only one ledger, he rushes from one set of data to another collecting together separate items to respond to highly varied requests for information. He is a godsend to management.

Executives often need information which spans departments or spans traditional boundaries in the corporation such as the engineering, accounting, personnel, production, and marketing functions. They need information on personnel implications of marketing decisions, or the impact on production of a new distribution strategy, or the labor costs associated with higher sales. Where each department has its own batch processing operations, the computer is of little value in answering such questions. However with a data-base approach the Dickensian super-clerk rushes from one department's books to another, searching and correlating the data. The structure of the data that are stored is agreed upon centrally so that interdepartmental usage is possible.

The data within a corporation (or other organization) will increasingly be regarded as a basic resource needed to run the corporation. As with other basic resources, professional management and organization of the data are needed. The importance of efficient use of data for production control, marketing, accounting, planning, and other functions will become so great in a computerized corporation that it will have a major effect on the growth and survival of corporations in a