

HEYDEN ADVANCES LIBRARY IN EDP MANAGEMENT

**Advances in
DATA BASE
MANAGEMENT**

Volume I

Editor: Thomas A. Rullo

HEYDEN

Advances in Data Base Management

VOLUME 1

Edited by
THOMAS A. RULLO



Philadelphia • London • Rheine

Heyden & Son Inc., 247 South 41st Street, Philadelphia, PA 19104, U.S.A.
Heyden & Son Ltd., Spectrum House, Hillview Gardens, London NW4 2JQ
Heyden & Son GmbH, Münsterstrasse 22, 4440 Rheine, Germany

© Heyden & Son Inc., 1980

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means electronic, mechanical, photocopying, recording or otherwise, without the prior permission of Heyden & Son Inc.

ISSN 0196-8718
ISBN 0 85501 602 7

PRINTED IN THE UNITED STATES OF AMERICA

Advances in Data Base Management

VOLUME 1

HEYDEN ADVANCES LIBRARY IN EDP MANAGEMENT

Edited by Thomas A. Rullo

Other Advances Series in the Library

ADVANCES IN DATA PROCESSING MANAGEMENT Volume 1

ADVANCES IN COMPUTER PROGRAMMING MANAGEMENT Volume 1

ADVANCES IN DISTRIBUTED PROCESSING MANAGEMENT Volume 1

ADVANCES IN DATA COMMUNICATIONS MANAGEMENT Volume 1

ADVANCES IN COMPUTER SECURITY MANAGEMENT Volume 1

LIST OF CONTRIBUTORS

- THOMAS R. FINNERAN, Corning Glass Works, Corning, New York 14830, U.S.A. (p. 54).
- MICHAEL HAMMER, Associate Professor of Computer Science, Massachusetts Institute of Technology Laboratory for Computer Service, Cambridge, Massachusetts 02139, U.S.A. (p. 172).
- HERMAN LAM, Database Systems Research and Development Center, Department of Computer and Information Sciences, University of Florida, Gainesville, Florida 32611, U.S.A. (p. 144).
- CHESTER C. LIU, Assistant Vice President, American Re-Insurance Company, New York, New York 10006, U.S.A. (p. 26).
- DENNIS MCLEOD, Professor, Computer Science Department, University of Southern California, Los Angeles, California 90007, U.S.A. (p. 172).
- STANLEY Y. W. SU, Professor, Database Systems Research and Development Center, Department of Computer and Information Sciences, University of Florida, Gainesville, Florida 32611, U.S.A. (p. 144).
- JACK F. THORNE, Ph.D., Professor, Department of Accounting and Finance, Auburn University, School of Business, Auburn, Alabama 36830, U.S.A. (p. 94).
- LARRY E. TOWNER, RCA Corporation, Mt. Laurel, New Jersey 08054, U.S.A. (p. 39).
- RONALD F. VOELL, Chief of Data Administration, General Dynamics Corporation, San Diego, California 92138, U.S.A. (p. 1).
- MYLES E. WALSH, Associate Director, Data Base Systems, CBS, Inc., New York, New York 10019, U.S.A. (p. 118).
- GABRIELLE WIORKOWSKI, Independent Data Processing Consultant, University of Texas at Dallas, Richardson, Texas 75080, U.S.A. (p. 70).
- JOHN J. WIORKOWSKI, Associate Professor of Statistics, University of Texas at Dallas, Richardson, Texas 75080, U.S.A. (p. 70).
- DAN ZATYKO, Zatyko Associates, Santa Ana, California 92705, U.S.A. (p. 194).

PREFACE TO THE HEYDEN ADVANCES LIBRARY IN EDP MANAGEMENT

During the past few years the rapid advances in EDP technology have been more than matched by a flood of published materials. It would be impossible to absorb all this new material and still be able to function in a working environment. Because of the information manager's plight, the HEYDEN ADVANCES LIBRARY IN EDP MANAGEMENT has been developed to provide a more useable information system.

A unique concept in the EDP information management field, the Library consists of six individual series, each dealing with a different area of information processing.

ADVANCES IN DATA PROCESSING MANAGEMENT
ADVANCES IN DATA BASE MANAGEMENT
ADVANCES IN COMPUTER PROGRAMMING MANAGEMENT
ADVANCES IN DISTRIBUTED PROCESSING MANAGEMENT
ADVANCES IN DATA COMMUNICATIONS MANAGEMENT
ADVANCES IN COMPUTER SECURITY MANAGEMENT

These series focus on the most current topics of interest across a broad spectrum. They are not, however, merely collections of papers or readings. Rather, each series presents chapters which have been selected with a specific information need in mind and developed by authors chosen for their expert knowledge and experience. This combination of breadth of material and depth of author knowledge results in a unique and concentrated information management program.

We intend to review the EDP information management field periodically and add to each series so that managers can gain insights into the latest developments. We will also be researching new areas of potential impact. The HEYDEN ADVANCES LIBRARY IN EDP MANAGEMENT is intended to be a continuing and expanding effort, and we would welcome any suggestions or guidance from our readers.

THOMAS A. RULLO
Editor

PREFACE

Data base management systems have evolved over the past 20 years from the complex early military systems developed to handle the massive amounts of data inherent in those applications to today's distributed systems, which service relatively unsophisticated needs. But while the technology has matured, the installation and operation of data base systems continue to pose as many problems as opportunities for the managers who must contend with them.

This book was developed to provide insight into those problems and thus, it is hoped, point the way to their solution while exposing the reader to some interesting opportunities to either improve or expand his use of data base management systems.

The authors contributing to this book represent a broad range of backgrounds and experience. They have attempted to address topics of current, as well as future, interest with the understanding that data base applications are never stagnant. The material deals with all stages of the life cycle of a data base management system, from the original planning and design through ongoing operation and future developments.

Ronald Voell's chapter, "Data Base Planning," reviews several planning methodologies and discusses a specific one in detail. In Chester Liu's chapter, "Creating the Proper Environment for DBMS," both the technical and political environments are addressed as he discusses proven techniques for ensuring that the DBMS is implemented in the proper fashion. To emphasize the point, he makes specific comparisons between a conventional and a data base environment.

In "The Hazards of Data Base Management Systems," Larry Towner draws upon his experience to outline potential hardware, software, legal, and people hazards associated with data base management systems. His down-to-earth approach and all-too-real examples make this valuable reading for both those considering DBMS and those currently operating in the environment.

Thomas Finneran's, "Data Base Design," provides a view of the alternatives available in the design process, from the feasibility study (where he emphasizes "business analysis") to the construction of a logical data base design, through both the design of the processing system and the ultimate tuning of the DBMS.

Gabrielle and John J. Wiorkowski discuss in detail both the technical and op-

erational benefits that accrue from the use of data base management systems in their chapter, "Evaluating the Benefits of Data Base Management."

The operational characteristics of data base management systems pose specific types of problems for auditors. In "Auditing Data Base Management Systems," Dr. Jack Thorne describes these characteristics and their implication for auditors. He makes specific recommendations to facilitate the implementation of controls and audit trails.

In "Data Dictionaries—A Pragmatic View," Myles E. Walsh views the latest developments in this field through the eyes of a user. His discussion of both the current trends and uses of data dictionaries includes practical insight and examples.

The impact of changing a data base on the application programs which access it is discussed in "Conversion of Application Programs to Account for Data Base Changes," by Stanley Y. W. Su and Herman Lam. Specific conversion techniques and alternatives are described.

In "The Semantic Data Model: A Conceptual Data Modelling Mechanism," Dr. Michael Hammer and Dr. Dennis McLeod discuss this natural application modelling mechanism, which can capture and express the structure of an application environment. They explain its use as a formal specification and documentation mechanism for a data base, as well as its ability to support a variety of powerful user interface facilities, and its use in the data base design process.

Dan Zatyko, in "Distributed Data Base," looks at the problems and potentials of distributed data bases by examining several of the most interesting approaches now in use.

As noted previously, the authors provide a broad spectrum of concepts through both their selection of material and their own specialized experience. Exposure to these concepts should prove beneficial, regardless of the stage of development of your data base organization.

THOMAS A. RULLO

CONTENTS

List of Contributors	xi
Preface to the Heyden Advances Library in EDP Management	xiii
Preface to Volume 1	xv

CHAPTER 1: Data Base Planning— Ronald F. Voell

Data Base Planning Prerequisites	2
A Background in Data Base Planning	3
Data Base Planning Methodology	6
<i>Study Plan</i>	
<i>Business Functions</i>	
<i>Processes of Business Functions</i>	
<i>Systems to Accomplish Processes</i>	
<i>Data Types</i>	
<i>Fact Finding</i>	
<i>Analysis Matrices</i>	
Spin-offs	10
General Dynamics Case Study	11
Monitoring the Plan	20
Future of Planning	22
Conclusion	23

CHAPTER 2: Creating the Proper Environment for Data Base Management Systems— Chester C. Liu

Climate in a Data Base Environment	26
Data Base Administrator	27
Data and Functional Analyses	30

DBMS Evaluation	32
Questionable Assumptions	35
<i>Benefit Analysis</i>	
<i>Conversion</i>	
<i>Data Redundancy</i>	
<i>DBMS Dependency</i>	
<i>Additional Mainframes</i>	
Summary	38
 CHAPTER 3: The Hazards of Data Base Management Systems— Larry E. Towner	
The Unseen Hazards	39
<i>Politics</i>	
<i>Laws</i>	
<i>Personnel</i>	
<i>Training</i>	
<i>Supporting Software</i>	
<i>Applications Software</i>	
<i>Hardware</i>	
Dollars and Cents	52
 CHAPTER 4: Data Base Design— Thomas R. Finneran	
Functional Business Analysis	54
Initiating the Logical Data Base Design	57
Analysis of Logical Views	59
Logical to Physical Data Base Design	61
Understanding the Data Base Management System	61
Data Structures Supported by a DBMS	61
Deriving a Physical Structure	62
Design Trade-offs	63
Technical Trade-offs	64
Choice of Access Approach	64
<i>Sequential Processing</i>	
<i>Indexed Approaches</i>	
<i>Direct Access</i>	
Best Access Methods?	67
<i>Blocking and Data Compression</i>	
<i>Data Placement</i>	
<i>Response Time and Throughput</i>	
Secondary Accessing	68
Security and Recovery	69
Pulling It All Together	69

**CHAPTER 5: Evaluating the Benefits of Data Base Management—
Gabrielle Wiorkowski and John J. Wiorkowski**

Integrated Systems	70
Minimize Data Redundancy	72
Data Integrity	73
Online DBMS	74
Data Independence	75
Backup and Recovery	79
Security	80
Performance	80
Cost/Benefit	84
Summary	87

**CHAPTER 6: Auditing Data Base Management Systems—
Jack F. Thorne**

Nature of DBMS	94
Audit Activities Affected by DBMS	97
Evaluation of Internal Control in DBMS	99
<i>Data Base Administrator</i>	
<i>Data Dictionary</i>	
<i>Controls of a Data Base System</i>	
Uses of Audit Software in Accessing Data from the Data Base	104
Methods for Accessing the Data Base	106
<i>Modify Generalized Audit Software to Access the Data Base Directly</i>	
<i>Modify Generalized Audit Software to Interface with DBMS</i>	
<i>Modify DBMS to Include Audit Functions</i>	
<i>Dump the Data Base</i>	
<i>Employ Utility Software</i>	
<i>Write Specialized Audit Software Packages</i>	
The Issue of Independence	108
Conclusion	110
Appendix A. EDP Accounting Controls	111
Appendix B. Computer Auditing Techniques	113

**CHAPTER 7: Data Dictionaries—A Pragmatic View—
Myles E. Walsh**

Concept and Definition	118
The Contents and the System	119
Use as a Documentation Tool	119
Relationship to a DBMS	123

System Features	126
<i>Determining Need</i>	
<i>Inputs</i>	
<i>Outputs</i>	
<i>Interaction with Other Software</i>	
<i>Security</i>	
Other Considerations	128
<i>Test and Production Status</i>	
<i>Data and Data Base Administration</i>	
Functions	130
A Contemporary Data Dictionary System: DD/D	131
Practical Application	133
Overview of Contemporary Data Dictionary Systems	134
<i>LEXICON</i>	
<i>UCCTEN</i>	
<i>DATA CATALOGUE</i>	
<i>DATAMANAGER</i>	
<i>INTEGRATED DATA DICTIONARY (IDD)</i>	
<i>CONTROL 2000</i>	
<i>CINCOM DATA DICTIONARY</i>	
Speculations on the Future of Data Dictionaries	138

CHAPTER 8: Conversion of Application Programs to Account for Data Base Changes—
Stanley Y. W. Su and Herman Lam

The Problem and Scope of Application Program Conversion. . .	146
<i>Conversion Within a DBMS Environment</i>	
<i>Conversion of the Host Language Program</i>	
Existing Tools and Research Work	149
<i>Existing Tools</i>	
<i>DML Substitution Technique</i>	
<i>Bridge Program Techniques</i>	
<i>Research Works</i>	
A General Methodology for Application Program	
Conversion	153
<i>Conversion at the Semantic Level</i>	
<i>Descriptive Model of an Application Program</i>	
<i>Conversion System</i>	
<i>Extraction and Preservation of an Application Program's</i>	
<i>Intended Operations</i>	
<i>Basic Access Patterns</i>	
<i>Data-Model-and-Schema-Dependent Representation</i>	
<i>DML/Host Language Templates</i>	

<i>Semantic Changes of Data Base and Transformation of Program's Access Patterns</i>	
<i>Semantic Transformations</i>	
<i>Transformations of Access Patterns</i>	
<i>Summary of the Methodology and Current Development Efforts</i>	
Conclusion and Future Work	165

CHAPTER 9: The Semantic Data Model: A Conceptual Data Modelling Mechanism—

Michael Hammer and Dennis McLeod

An Overview of the Semantic Data Model	173
<i>The Basic Structure of an SDM Data Base:</i>	
<i>Classes and Attributes</i>	
Modelling with the SDM	183
<i>Problems with Conventional Data Models</i>	
<i>Related Work on Semantic Modelling</i>	
<i>Redundancy</i>	
<i>Modelling Associations in the SDM</i>	
Applications of the SDM	188
Summary	190
Appendix: A Portion of the SDM Schema for the TMAE	191

CHAPTER 10: Distributed Data Bases—

Dan Zatyko

Decision Support Systems	194
Performance-Related Distributed Data	196
Host Mirror (Hierarchically Structured Satellite Files) ...	197
Host-Satellite Data Base Relationships	198
IBM's Distributed Systems Executive (DSX)	201
Host-Connected Case Study:	
Food Processing Application Network	201
Performance-Related Distributed Data Base Approaches .	203
Directions of Standards Organizations	206
Conclusions	207

Chapter 1

DATA BASE PLANNING

Ronald F. Voell

General Dynamics, Data Systems Services, San Diego, California 92138

Management can be defined as a combination of activities which include forecasting, planning, organizing, and controlling. The combination is symbiotic. Overlooking one activity renders the others ineffective. Control is a vague concept in the absence of forecasting, planning, or organizing. Planning is, therefore, a necessary part of controlling and of managing. What can then be said about the state of affairs that is more the rule than the exception in data processing today? Is a significant organizational resource being managed in the absence of a plan for data base definition and development? The answer to these questions depends on the answer to a bigger question. Why should data bases be managed?

We live in a point in history where many merging revolutions are occurring. Rapid social and technological change is accepted as given. In this environment, forecasting and planning are taking on increased importance. The wisdom of predicting the future is accepted as proven even though studies about it lack perfection. Planning for that future is also prudent. Planning is a process which seeks to provide a roadmap either to make a forecast a reality or to change conditions so a forecast can be modified.

Today, one of the most powerful forces in the world is technology. Change in this area is combining with changes in other areas and producing a profound effect on our lives. The automated processing of information is no exception. Data processing is both a source of and a reaction to change. Forecasts predict continuing increases in the hardware performance/cost ratio and development of more effective software. The growth is being driven partly by increased labor costs. This trend has been observed and commented on by many authorities.⁴¹ Data base and data administration are technological and managerial reactions to the fact that hardware and labor are exchanging their positions of relative economic importance in the overall cost of data processing.

But the data base and data administration have, by and large, not lived up to the expectations and promises made in most writings. The corporate data base integrated to reduce redundancy, accessible to multiple applications, protected from the unauthorized, and controlled so that it contains only valid information is a myth. Today data base is largely just another access method. Similarly, the data ad-

ministration function is not controlling a global view of information. The real promise and benefits of the controlled corporate data base are valid, but are still to be realized.

The lack of a plan creates problems in multidivision organizations. The managements in these types of business are questioning a significant investment in application systems and, in an effort to reduce future costs, are more and more attempting to achieve common systems. They reason that if a system could be built once and used in all company divisions, then costs could be reduced. But interfacing the common system into the differing data structures that exist across divisions generally means that the common systems are not economically feasible.

The traditional way of designing systems and data bases limits their scope to the organizational boundaries. These systems basically computerize clerk tasks. The system user participates in the design process by way of defining the reports and the inputs to the system. Indirectly they also define the data. The needs are expressed with a departmental perspective.

Integrating systems and the data bases that evolve on this basis is an impossible task. No clear picture exists of how an application system fits into future needs. There is no view of where the system dovetails into the overall data structure. Zachman states the problem:

If the information systems architecture has not been defined, then when the projects are translated into an operational status, they do not build anything. They merely exist on a stand-alone basis. They do not relate to each other. They do not relate to the flow of information within the business. They do not relate to the relationships between the structural entities of the business. They relate only to an isolated problem identified within the business, and their value is only the value of solving that specific problem.⁵⁷

All of this background suggests the simple question: Why is this situation permitted to continue? If the technology is available and if goals are known, why have they not been achieved? The biggest handicap is lack of understanding of where systems development is heading over the long haul. No roadmap to the future exists; therefore, there is nothing with which to compare individual application needs as requirements for them are identified. Systems are developed by a random process with little regard to possible interrelationships with other systems, most of which may be future developments. To say it another way, we react; we don't plan.

This problem will not be solved overnight; after all, the problem was not created overnight. But without a plan, the present course will continue, building a bigger and bigger future problem. This chapter addresses a methodology for developing such a plan. It is a plan which can only be slowly implemented.

DATA BASE PLANNING PREREQUISITES

Planning is not the first step in the management process. Forecasting, defining objectives, fact finding, and installing the tools for collecting and analyzing a large

volume of information are important prerequisites for effective planning.

A plan is a roadmap to the future. It goes without saying that some things ought to be known about the future before planning can be undertaken. Every company's forecast tends to be unique. The future needs to be relatively clear, at least over the period of time that the plan will be implemented. The fact that forecasting is an inaccurate science must not be used as a justification for completely overlooking this important step. Technological, organizational, and business predictions have an important impact on data base plans, even though the predictions lack perfection.⁴

There are other prerequisites to data base planning. A most important one is the support from the organization's top management and the communication of that support to all levels of the organization. Moreover, long-range data base planning involves the collection of considerable amounts of information. The collection of this information has implications of organizational change. Problems often occur with data processing people themselves. Integrated information means a reduced programming staff. Personal concerns of people can reduce the effectiveness of studies. It is therefore extremely important that top management clearly and perhaps forcefully communicate its goals through the organization. Management should not expect the data processing department to relay those objectives for it.

Another important prerequisite is a mature organization which has an understanding that data is a resource and which has taken steps to get data under management control. A specific symptom of the organization's appreciation of "data as a resource" is the existence of a data administration function which has centralized standards and control over data.

An additional prerequisite is the installation of a data dictionary which has a significant percentage of the organization's management systems documented in it. Data base planning involves collecting and analyzing large quantities of data. It is not practical to accomplish this task manually. Some experts project that it requires somewhere between 5,000 and 15,000 discrete elements of data to run a company. The study of the company will show each element occurring several times. By remembering that a corporate data base is being built, one can see that the data are again multiplied by the number of divisions. Without a data dictionary, the project gets bogged down with detail.

A BACKGROUND IN DATA BASE PLANNING

Data base planning is only one part of the total planning effort that is conducted in the organization. As such, it affects and is in turn affected by much of this other planning. Data base planning is, therefore, not done in a vacuum. It is closely related to application systems planning and, to an extent, organizational planning. In this chapter, these three plans will be referred to as the system architecture plan. Although the thrust of this writing will be on data base planning, it is important