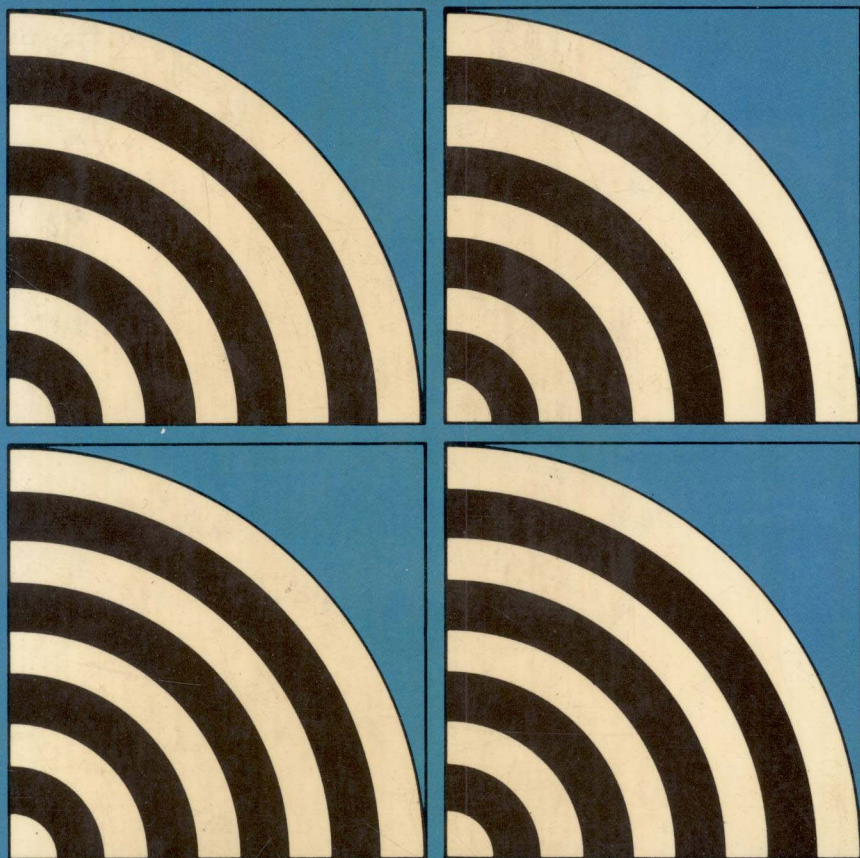




British Computer Society
Monographs in Informatics

Database Design: A Classified and Annotated Bibliography

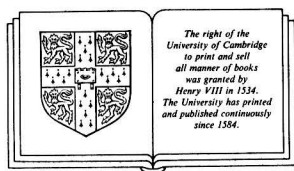
Maristella Agosti



Database design: a classified and annotated bibliography

MARISTELLA AGOSTI

*Facolta' di Scienze Statistiche
Universita' di Padova*



CAMBRIDGE UNIVERSITY PRESS

Cambridge

London New York New Rochelle

Melbourne Sydney

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
10 Oakleigh Road, Melbourne 3166, Australia

© Cambridge University Press 1986

First published 1986

Printed in Great Britain at the
University Press, Cambridge

British Library cataloguing in publication data

Agosti, Maristella

Database design: a classified and annotated
bibliography.—(The British Computer Society
monographs in informatics)

1. Database management—Bibliography 2. File
organization (Computer Science)—Bibliography

I. Title II. Series

016.00574 Z5643.D36

Library of Congress cataloguing data available

ISBN 0 521 31123 3

FOREWORD

A rapidly growing area of application of information technology is the use of databases. There is a large and ever-increasing literature, there are proprietary packages available by the hundred all making claims to solve the world's problems. The non-specialist user is confused and is not even sure what questions should be asked to decide on how to establish a system that meets his needs. Dr Agosti's carefully annotated bibliography of literature on the subject up to the end of 1984 will be of considerable help to all, specialist and non-specialist alike. It forms a welcome addition to this series of monographs.

P.A. Samet
Editor,
British Computer Society,
Monographs in Informatics

CONTENTS

<u>Foreword</u>	2
<u>Preface</u>	3
<u>Abbreviations and acronyms</u>	9
List of subjects	11
Subjects directory of alphabetically-ordered references	13
References in alphabetical order	23
<u>Index of authors</u>	89

FOREWORD

A rapidly growing area of application of information technology is the use of databases. There is a large and ever-increasing literature, there are proprietary packages available by the hundred all making claims to solve the world's problems. The non-specialist user is confused and is not even sure what questions should be asked to decide on how to establish a system that meets his needs. Dr Agosti's carefully annotated bibliography of literature on the subject up to the end of 1984 will be of considerable help to all, specialist and non-specialist alike. It forms a welcome addition to this series of monographs.

P.A. Samet
Editor,
British Computer Society,
Monographs in Informatics

PREFACE

1.0 INTRODUCTION

Database design is a process concerned with the construction of a database as a subsystem of an information system. Thus the database design process is an integral component of the information system design process. If the decision has been made to develop an application using database technology, the database design process can be started. Different related applications can be arranged to share the same database. The next section introduces the database design subject and the related terminology used to define a subject directory. In fact a list of subjects has been used as a directory to form a cross-reference to the pertinent documents. The references appear alphabetically later in the bibliography. Terms that have been defined by standard organisations or by other professional associations are adopted with the same meaning those bodies have ascribed to them; relevant references are (ISO, 1981), (ISO, 1982), and (SECT/IEEE, 1982).

Each reference includes the surname of the author(s) and the year of publication, for example, (Adiba, 1980) or (Adiba and Delobel, 1977). When the reference includes three or more authors, it reads (Adiba et al, 1976). The list of references is in alphabetical order by surname of the first author. In the case of multiple listings by a single author, the earliest publication appears first. When different references are given to an author alone or the author with others, all the single-author listings precede the multiple-author listings (e.g. (Chen, 1976), (Chen, 1982), (Chen et al, 1982)).

The abbreviations used are reported at the end of the preface. The abbreviations are not used in the reference title which is always reported as it is in the original text.

The accents have been omitted from Italian and French words.

2.0 DISCUSSION

The database design process is used to transform and organise unstructured information and processing requirements concerning the application, through different intermediate representations, to a complex representation which defines schemas and functional specifications. Various documents which record the intermediate representations and the meanings of all the pertinent classes of objects are produced during this process. The database which is constructed during this process is used in the enterprise for different purposes and by diverse categories of users. Parts of the information system which are supported by the computer are often referred to as applications.

The design process is usually divided into components which produce intermediate representations. The division into components is dependent upon the present knowledge of the process and the present technology. The divisions are useful for managing the process and permitting the exchange of information and intermediate results between the staff involved in the design. In the ensuing paragraphs the underlined terms correspond with the headings in the bibliography.

The data stored in a database system is used essentially to make decisions at the different levels of the enterprise. The stored data concerns the information system objects and their attributes; the meaning the users tend to ascribe to the data can change in the different departments of the enterprise and on the use they make of it. For these reasons a major exercise has to be carried out in parallel with the database design process. This is the construction of a holder for the meanings ascribed by the designer to the data held in the database. The data descriptions are called metadata; the holder of them is called a data dictionary and the system that manages it a data dictionary system.

If it is shown in the preparatory cost/benefit analysis of the application, that it is advisable to manage the data with a database management system, then it will be economical to make a substantial investment in the design of the database, so that it can be used by diverse applications. Errors made during the design process can affect the application's entire life. The process therefore is of great importance to the enterprise and it is necessary to pay much

attention to it.

The documents referring to the interface problems with the information system area and other areas related to the database design are grouped under the database design 'environment' heading. A subsequent section is devoted to the system development processes (prototyping, ...); this section refers to documents pertinent to this general topic.

The documents pertinent to information system design methodologies are grouped under the heading information systems design methodologies and evaluation.

So far a common consensus has not been reached on the components the database design process must have. Most of the authors agree on a division of the process into four or five components (Lum et al, 1979), (Teorey and Fry, 1982) and (Yao et al, 1982b); a division of the process into four components is adopted in this bibliography:

- 1 Information requirements design;
- 2 Conceptual database design;
- 3 Implementation design;
- 4 Physical design.

The information requirements design is the first component of the application design. It is the interface between the analysis process and the design process. It represents the process of mapping analysis into design. It leads to the specification of the design requirements for the pertinent part of the information system to meet the information requirements of the enterprise.

The conceptual database design component leads to the construction of the conceptual schema which provides a unique central description of the various information contents that may be in the database. The conceptual schema is a DBMS-independent information structure which is obtained through the consolidation of the user information requirements specifications. Both the static (or passive) and the dynamic (or active) characteristics of the data are taken into consideration.

The implementation design incorporates the conceptual model into the structure of the selected database management system and analyses the possible transactions. The file design starts in this component together with the access paths study.

The physical design component serves to construct the storage schema. The physical space for records and indices and the physical media that are to be utilised are defined.

The logical database design term has not been used to name one design component, because it has so many different meanings in the literature and its use can only cause confusion. Also conceptual design can be a misleading term because it is applied to different practices; the conceptual design component is meant to embrace the integration of all the concepts which are necessary to support the various application views of data; the conceptual design produces a conceptual view of data in which the particularities of specific views of data are resolved (Smith et al, 1982). The documents that deal with the conceptual design at a very general level are grouped under conceptual modelling.

The meaning ascribed to data design is that introduced by Wasserman and Botnick in (Wasserman and Botnick, 1981). That is a wide range of topics on data structures and database design. In particular the topics across the programming languages/database management interface and the design techniques that propose systematic methods for the design activity.

Data analysis refers to the techniques which study the nature of the data pertinent to the application.

Different categories of users have different targets in their work and they tend to see the data from different perspectives or views; a view is the representation of a portion of the application and it is the part that a group of users needs to deal with most often for data management. A snapshot of a selected portion of the database allows some applications to view the database 'as of' a specific time without having to execute a query at that specific time (Adiba and Lindsay, 1980). The problem of mapping is referred to in the appropriate section.

A data model is a way of representing data and its interrelationships. A data model describes the data of the pertinent part of the information system, because it can be used as a tool for capturing the meaning of data as related to the complete meaning of the information system. Diverse data models are usually used in the different components of the database design process to describe the data of the pertinent part of the application. In the conceptual database design component a semantic data model is adopted, where semantic data model identifies all the data models that have been constructed to explicitly capture the semantics of the data

with their constructs. The semantic data model derives its name from the fact that it is used during the conceptual design to capture the meaning, or semantics, of the application. While a processable data model is used in the implementation design component. It is called processable because the data representations, or schema, are processable by the computer. Thus the data models play a major role in the database design process and a section has been devoted to the subject.

The database design methodologies, techniques, and methods which are practiced today are part of the development techniques for computer-based systems. Their development is directly effected by the developments taking place in related areas such as programming languages and hardware technology. A "database design methodology" is an integrated collection of methods and techniques, which supports the complete database design process. A "technique" in this design context provides a systematic way of doing a part of the design. When a technique is applied correctly, it leads to a foreseeable result. A technique does not fulfil the requirements of integration and completeness that are required of a methodology. A "method" is an organised set of ideas that are used in doing a specific activity. It is a formal way of doing the design of only an activity. Several database design methods, techniques and methodologies have been developed and proposed in literature; the database design methodologies, methods and techniques section indexes pertinent documents. The design of new database architectures section addresses new design approaches.

The distributed database design section refers to papers which are focused on the distribution of the application.

The section on reference models for the database design process refers to papers which are concerned with the construction or the use of a general framework of the database design process.

The database design process is complex and many different aspects need to be addressed by the designer. The keeping of the design documentation consistent with the design development and modification requires a big effort. The database restructuring, reorganisation and conversion section indexes documents on the different aspects of modification of a database. Computer aided database design tools are being developed to support the designer in one or more components of the design process; pertinent papers are referred to in the computer aided database design section.

Relevant standards and ANSI/X3/SPARC, the last two terms of the directory, refer to relevant activities of bodies concerned with international standards.

ACKNOWLEDGEMENTS

The research necessary for the preparation of this bibliography would not been possible without the effective help of the librarians of the Main Library of the Thames Polytechnic, London, to whom sincere thanks are due.

ABBREVIATIONS AND ACRONYMS

ACM	Assoc. for Computing Machinery
AICA	Assoc. It. di Informatica e di Calcolo Automatico
Ann.	Annual
Archit.	Architecture
Assoc.	Association, Associazione
BCS	British Computer Society
Bull.	Bulletin
CADD	Computer Aided Database Design
Chap.	Chapter
Co.	Company
Comm.	Communications
Conf.	Conference
CREST	EEC Scientific and Technical Research Committee
DB	Database(s)
DBMS	DB Management System(s)
DBS	Database System
DDB	Distributed Database(s)
DDBMS	Distributed DBMS(s)
DD/D	Data Dictionary/Directory systems
DDL	Data Definition Language
DDS	Data Dictionary System
DDSWP	Data Dictionary Systems Working Party
DML	Data Manipulation Language
Dept.	Department
Doc.	Document
Ed(s)	Editor(s), Edition
Engin.	Engineering, Engineer(s)
E/R	Entity-Relationship
IEEE	The Institute of Electrical and Electronics Engin.
IFIP	International Federation for Information Processing
Inform.	Information
Int.	International
IS	Information System(s)
ISO	Int. Organization for Standardization
It.	Italiano, Italiana, Italian
Jour.	Journal
Lab.	Laboratory
Ltd.	Limited
MIS	Management Information Systems
Nat.	National

NBS	United States (US) National Bureau of Standards
No.	Number
NYU	New York University
OSI	Open Systems Interconnection
p(p).	page(s)
Prin.	Principles
Proc.	Proceedings
Pub.	Publication(s), Publisher or Publishing
RDBMS	Relational DBMS
Rep.	Report
Res.	Research
Ser.	Series
SIGMOD	Special Interest Group on Management of Data
Surv.	Survey(s)
Symp.	Symposium
Syst.	System(s)
TC	Technical Committee
Tech.	Technical, Technology
Trans.	Transactions
Univ.	University
VLDB	Very Large Data Bases
Vol.	Volume
WG	Working Group

Short form of each month name.

The abbreviation of each state of the USA used with zip code.

LIST OF SUBJECTS

GENERAL

- Database design 'environment'
- System development processes (prototyping, ...)
- Conceptual modelling
- Data design
- Data models
- Data analysis

COMPONENTS OF THE DATABASE DESIGN PROCESS

- Information requirements design
- Conceptual database design
- Implementation design
- Physical design
- Design versions

OTHER ASPECTS OF THE DESIGN

- Mapping
- Snapshots
- Views

DATA DICTIONARY SYSTEMS

- Data dictionary systems

METHODOLOGIES AND REFERENCE MODELS

- Information systems design methodologies and evolution
- Database design methodologies, methods and techniques

Reference models for the database design process
Design of new database architectures

MANAGEMENT OF THE DESIGN

Computer aided database design
Database restructuring, reorganisation and conversion

DISTRIBUTED DESIGN

Distributed database design

STANDARDS

Relevant standards
ANSI/X3/SPARC

SUBJECT DIRECTORY OF REFERENCES

GENERAL

Database design 'environment'

- (Amble et al, 1979)
- (Atre, 1980)
- (Baker and Holloway, 1984)
- (Bhabuta, 1984)
- (Bubenko et al, 1978)
- (Ceri, 1980)
- (Champine et al, 1980)
- (Chen and Yao, 1977)
- (Cookson, 1983)
- (Couger et al, 1982)
- (Date, 1981a)
- (DDBWG/BCS, 1983)
- (Deen and Hammersley, 1980)
- (Fry and Teorey, 1978)
- (GAO, 1979)
- (Hainaut, 1981)
- (Inmon, 1981)
- (Kambayashi et al, 1982)
- (Kent, 1981b)
- (McCracken and Jackson, 1982)
- (Maggiolini, 1981)
- (Mumford, 1984)
- (Nijssen, 1976b)
- (Olle et al, 1982)
- (Stamper, 1977)
- (Sundgren, 1975)
- (Sundgren, 1978)
- (Teorey et al, 1982)
- (Ullman, 1980)
- (Wasserman et al, 1981)
- (Weber and Wasserman, 1979)
- (Wiederhold, 1977)
- (Wood-Harper and Fitzgerald, 1982)
- (Yao et al, 1982a)
- (Yao and Kunii, 1982)