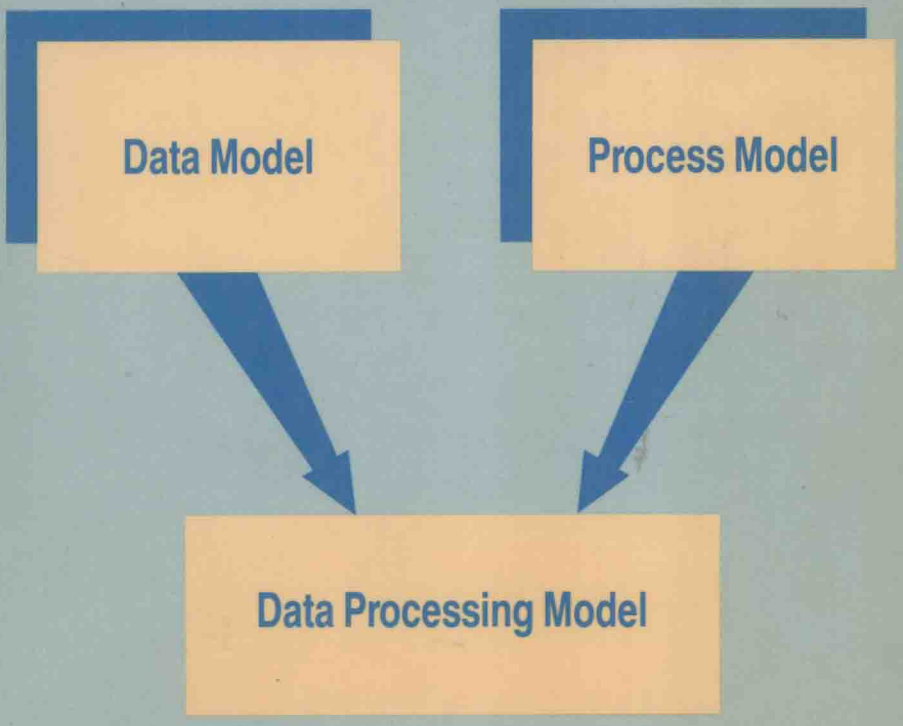
Information and Data Modelling

Information Systems Series

BLACKWELL SCIENTIFIC PUBLICATIONS

David Benyon



INFORMATION SYSTEMS SERIES

INFORMATION AND DATA MODELLING

DAVID BENYON BA, MSc
Department of Computing
The Open University

BLACKWELL SCIENTIFIC PUBLICATIONS
OXFORD LONDON EDINBURGH
BOSTON MELBOURNE

Blackwell Scientific Publications
Editorial offices:
Osney Mead, Oxford OX2 0EL
(*Orders*: Tel: 0865 240201)
8 John Street, London WC1N 2ES
23 Ainslie Place, Edinburgh EH3 6AJ
3 Cambridge Center, Suite 208
Cambridge, Massachusetts 02142, USA
107 Barry Street, Carlton
Victoria 3053, Australia

First published 1990

Set by DP Photosetting, Aylesbury, Bucks
Printed in Great Britain by
St Edmundsbury Press Ltd
Bury St Edmunds, Suffolk.

DISTRIBUTORS

Marston Book Services Ltd
PO Box 87
Oxford OX2 0DT
(*Orders*: Tel: 0865 791155
Fax: 0865 791927
Telex: 837515)

USA
Publishers' Business Services
PO Box 447
Brookline Village
Massachusetts 02147
(*Orders*: Tel: (617) 524 7678)

Canada
Oxford University Press
70 Wynford Drive
Don Mills
Ontario M3C 1J9
(*Orders*: Tel: (416) 441-2941)

Australia
Blackwell Scientific Publications
(Australia) Pty Ltd
107 Barry Street
Carlton, Victoria 3053
(*Orders*: Tel: (03) 347 0300)

British Library
Cataloguing in Publication Data
Benyon, David
Information and data modelling
(Information systems series)
1. Databases. Management
I. Title II. Series
005.74

ISBN 0–632–01625–6

Library of Congress
Cataloging-in-Publication Data
Benyon, David.
Information and data modelling/David Benyon.
p. cm. — (Information systems series)
Includes index.
ISBN 0–632–01625–6
1. Data base design. 2. Data base management. I. Title.
II. Series: Information systems series
(Oxford, England)
QA76.9.D26B46 1989
005.74–dc19 88–34657
CIP

INFORMATION SYSTEMS SERIES

Consulting Editors

D. E. AVISON
BA, MSc, FBCS
Department of Computer Science and Applied Mathematics, Aston University, Birmingham, UK

G. FITZGERALD
BA, MSc, MBCS
Oxford Institute of Information Management Templeton College, Oxford, UK

This is a brand new series of student texts covering a wide variety of topics relating to information systems. It is designed to fulfil the needs of the growing number of courses on, and interest in, computing and information systems which do not focus purely on the technological aspects, but seek to relate these to business or organizational context.

For my Mother and Father,
Charlotte Louise and Joseph Donald Benyon

Foreword

The Blackwell Scientific Publications series on Information Systems is a brand new series of student texts covering a wide variety of topics relating to information systems. It is designed to fulfil the needs of the growing number of courses on, and interest in, computing and information systems which do not focus purely on the technological aspects, but seek to relate these to business or organisational context.

Information systems have been defined as the effective design, delivery, use and impact of information technology in organisations and society. Utilising this fairly wide definition, it is clear that the subject area is somewhat interdisciplinary. Thus the series seeks to integrate technological disciplines with management and other disciplines, for example, psychology and philosophy. It is felt that currently these areas do not have a natural home, they are rarely represented by single departments in polytechnics and universities, and to put such books into a purely computer science or management series restricts potential readership and the benefits that such texts can provide. This series on information systems now provides such a home.

The books will be mainly student texts, although certain topics may be dealt with at a deeper, more research orientated level.

The series is expected to include the following areas, although this is not an exhaustive list: information systems development methodologies, office information systems, management information systems, decision support systems, information modelling and databases, systems theory, human aspects and the human-computer interface, application systems, technology strategy, planning and control, and expert systems, knowledge acquisition and representation.

Whereas the first text in the series, *Information Systems Development: Methodologies, Techniques and Tools* (D. E. Avision and G. Fitzgerald), looked at many of these areas in overview form, the second text in the series by David Benyon, *Information and Data Modelling*, concerns itself with one very important aspect of the information systems world, the world of data, in some depth. The management of data is a key issue as it enables information

to be made available to information systems users. The book focuses on how to design the database which will be used by the information system so that it accurately and comprehensively represents the enterprise which is to be the subject of the information system. *Information and Data Modelling* also deals with the vital role which the data administrator has in developing and implementing an information system and the importance of data related aspects in the strategic planning for information systems.

David Avison and Guy Fitzgerald
Consulting Editors
Information Systems Series

Preface

Information and data are two invaluable resources in our society. To fully exploit their potential they need to be managed. This book describes the importance of these resources, how they are related and why the design of the database is vital to the successful production of information. However, information cannot be managed as easily as other resources such as finance, energy, materials or personnel. The costs and benefits of information are still poorly understood and difficult, if not impossible, to quantify. Perhaps one should say 'currently poorly understood' because our knowledge of information and its impact is growing daily.

The fact remains that we have yet to begin to widen our horizons of what is possible given the impact of new computer technologies. Expressions such as 'information revolution' or information society', mean about as much now as 'industrial revolution' meant to the farming communities of the early nineteenth century. We can anticipate that we will be able to perform the same tasks more quickly and more cheaply, but we cannot predict the fundamental changes in the structure of business organisation and society which may or may not occur as a result of this revolution.

Information is part of the infrastructure of society. It can be of interest in itself, but its real impact is to facilitate new activities or to transform current practices. The use of terms such as 'paperless office' is reminiscent of terms such as 'horseless carriage' or 'wireless telegraph'. These expressions try to describe wholly new phenomena (motor cars and radio) in terms of existing structures (horses and telegraphs). The real impact of such inventions and their eventual form frequently have little resemblance to their forebears. Whether the 'paperless office' will be a fundamentally different entity from current offices, requiring different skills and changing our concepts of work and business management, remains to be seen.

It is, therefore, with some trepidation that I commence a book about something which is still in its infancy and which may have an impact far beyond our current perceptions. But electronic computers have been around for 40 years; surely we have learnt enough to anticipate their impact? Whilst it is true that we have come a long way in 40 years, the current 'generation' of

computer hardware and software (at the time of writing we are somewhere in the middle of the fourth generation) has significantly altered our view of what computers can do. The developments in micro-electronic technology have been quite dramatic and future enhancements will increase speed and power another ten-fold in the next decade. This quantative change in power results in a qualitative change in the product. Coupled with this are the developments in electronic communications via satellites, radio and fibre optics and novel computer architectures, which open up many more alternatives for the way in which processing is accomplished. On the software side, our understanding of the methods of organising and accessing data, and the provision of increasingly powerful methods for communicating between the human and computer, have followed swiftly in the wake of the hardware developments.

As a result of these (continuing) changes, computers can now be applied to tasks which were once (perhaps only ten years ago) seen as sitting firmly in the domain of people. Leaving aside the potential associated with the development of 'intelligent' computers (the fifth generation), the type, speed, reliability and quantity of information which can be made available through fourth-generation hardware and software, and the range of applications to which computers can now be put is changing the way organisations can be managed. One of the most important of these developments has been the emergence of powerful software called database management systems (DBMS) and the consequent shift of attention to the shared pool of data – the database.

In many ways this book deals with only half the problem. It concentrates on how information can be produced rather than what information is needed to run an organisation. This is inevitable at present as we have yet to experience the full impact of information revolution. Many business managers have been doing their job with poor quality information at their disposal. Much of their time has been spent checking that things have been done properly or dealing with day-to-day operational problems. With well-managed information at their disposal, the nature of managers' jobs is changing and with this change will come demands for a different sort of information and facilities to investigate, analyse and manipulate the data which is used to represent their organisation.

Information is made available through a collection of objects and activities which we call an information system. At the kernel of an information system is the database which is the 'warehouse' for the data needed to produce information. The database is frequently maintained on a computer, and a more pedantic description is thus 'computer-based information system'. One

of the most difficult problems facing the information system designer is how to organise the data so that information can be obtained. This is the task of database design culminating in the production of a **data model**. This is a representation of the organisation constructed from data and orientated towards the information requirements of managers.

The book is deliberately non-technical and the reader requires no knowledge of computers or computing. The absence of a technical treatment is important for several reasons:

(1) Most other books offer technical details and these are referenced at the end of each chapter of this text. The reader requiring more depth can follow up his/her interests as they wish.

(2) It allows the text to concentrate on issues which are independent of implementation. The arguments for the design of databases are as relevant for a completely manual system as they are for a computerised system.

(3) The internal workings of computers are irrelevant to most people who want to use computers for a specified purpose. We can make telephone calls without understanding a switchboard and drive cars without understanding the workings of a gearbox. This is not to say that such knowledge is uninteresting in itself, nor that more knowledge may enable the telephone or car user to exploit the facilities more efficiently. It simply means that such detailed knowledge is not necessary to make these things work.

The non-technical orientation and management emphasis of the text makes it suitable for a wide variety of readers:

(1) The business (or other enterprise) manager who wants to exploit the opportunities offered by computers must be able to analyse and specify his/her requirements for information.

(2) The business or management studies student will frequently undertake a course in computing and/or information systems. This book provides an ideal set text for such a course.

(3) The computer studies student may do a detailed course on database systems, but will rarely cover the semantic issues emphasised here.

(4) Students and practitioners in a variety of disciplines will find that the techniques for understanding and structuring data described in this book are useful for understanding a whole range of problems, particularly if they are to use computers in the solution to those problems.

This book pulls together many of the most recent developments in database and information systems theory and presents a comphensive and

coherent approach to database design. It is suitable at postgraduate and undergraduate levels, and where it does not cover a topic in complete detail, annotated references are given to relevant papers and chapters of books. The first part of the book (Chapters 1–3) is concerned with understanding information, data and the management of data. Chapter 4 gives a general introduction to the principle of modelling and the data models which are required in order to represent the enterprise under scrutiny. Chapter 5 introduces the principles of the **relational data model** and Chapter 6 covers how relations should be designed. In Chapter 7, a complementary data model – the **entity-relationship model** – is described in detail. These models are brought together into a practical methodology for data modelling in Chapter 8. The final chapter looks at the important area of strategic information modelling.

ACKNOWLEDGEMENTS

Over the years, many people have helped me to understand computers, information and information systems and have, therefore, contributed indirectly to this book. In particular I would like to thank Rona Bassett and Bill Lamey of Osprey Computer Services for starting me on the road of commercial computing and instilling the principles of good file, program and interface design. I would also like to thank Professors Conway and Parkin of Leicester Polytechnic for their faith, help and encouragement throughout my stay there.

I also owe many a debt to my other colleagues at Leicester Polytechnic, in particular David Howe for his lucid introductions to data analysis and Steve Skidmore with whom I have had many a stimulating discussion. My thanks to him also for the basis of the Disco Jones case study. Gillian Mills, Peter Messer and Ray Farmer all commented on earlier drafts of the manuscript.

I would like to thank Bernie Glasson of Curtin University, Perth, Australia who provided me with a valuable opportunity to spend time researching the book. Many of my colleagues during my stay at Curtin University, particularly Peter Robinson, contributed to the book, albeit unknowingly. My thanks also to Mike Newton, Hugh Robinson, Ray Weedon and Mike Field at the Open University with whom I have been working on the course entitled *Data Models and Databases*. This experience has further shaped my views on database systems. Responsibility for the final content, however, is mine alone.

Finally I would like to acknowledge all help that I have received from the technical and secretarial staff at both Leicester Polytechnic and the Open University, particularly Valerie Kirby.

David Benyon

Contents

Chapter 1
Information

1.1 INTRODUCTION

Every day we use information to enable us to deal with the world. Without accurate, up-to-date information, we make mistakes and misjudgements. For example, you arrive at the bank after it has closed because you did not know the correct time, or arrive at the station as the train disappears because you had an old train timetable. Information is such a familiar concept that we rarely think about it. But what is information? Where does it come from? How can we produce it reliably, accurately and on time? How much information do we need and for what purposes?

This text is concerned with information for management. Information which will enable people to plan, control and monitor their organisation whatever the size or objectives of that enterprise; as machines need electricity to function, so managers need information. With the correct sort of information delivered in the right place at the right time, management will be in a position to take actions appropriate to the situation.

The terms 'management', 'manager', 'managing' and so on are used in this text in a very broad sense. The use of these words does not imply any particular size or type of organisation. Similarly, the terms 'organisation' and 'enterprise' are used to mean any purposeful activity – be that a large profit-making business, a public utility or an individual's hobby. The reason for stressing that the purpose of the information is to manage an enterprise is important because of the structure of such information.

Most writers on the subject accept the importance of information. Tully, writing in the special issue of the *Computer Journal*, refers to information as '... essential to render human activity effective' (p.208) and Scarrott in the same edition argues that the function of information is to '... control the actions of the components of an organised system' (p.203). Langefors and Samuelson state that '... the first analysis of every organisation's requirements ... shows that it needs some information system in order to exist' (p.1).

The assumption underlying this last statement is that the world can usefully be perceived as a multitude of systems. A system is a coherent set of

interdependent components which exists for some purpose, has some stability and can usefully be viewed as a whole. The class of systems which are dealt with in this text have been called 'human activity systems' by Checkland. They consist of people, norms, conventions and, optionally, artefacts, designed to serve human needs. The National Health Service is a system; a football match is a system; Marks and Spencer is a system; my family is a system.

Each of these also has an **information system**, the purpose of which is to manage (that is, plan, control, monitor and maintain) the human activity system. The information system may be informal (for example, casual conversation), or it may be formalised to a greater or lesser extent. For example, the information system for my family is totally informal. Some of the football match is formalised through the use of flags and whistles, but much remains informal as the referee has a 'quiet word' with an offending player, rather than 'taking his name' by formally recording it in his book. The

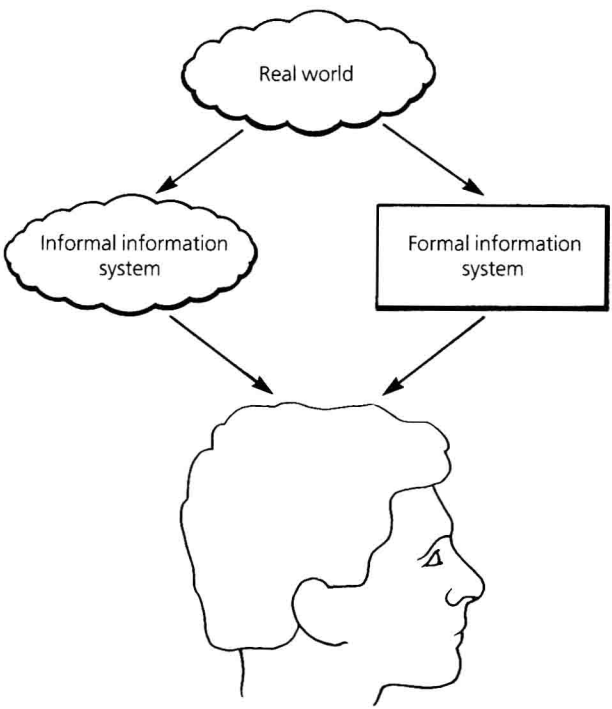


Fig. 1.1. Formal and informal information systems

National Health Service and Marks and Spencer also use informal information systems, but much more of their information comes from formal information systems which use computers. It is primarily the formal information systems, particularly those employing computers, which are the concern of this text. A diagram of information systems (adapted from Land) is shown in Figure 1.1.

It is not the purpose of this text to tell managers what information they need to control their organisations. The book covers how to approach the gathering and structuring of data to enable the production of information. It highlights methods for the analysis of information requirements and describes techniques for information systems design. However, each manager will need a different sort of information suitable for the purpose at hand. Too much information can be unhelpful and confuse an issue and information which changes too rapidly can be unsettling and counter-productive. Considerable thought should be given by managers to the sort of information which is required.

It can be difficult, particularly for those unfamiliar with computers, to decide what information is required for effective management. The ability of the computer to deliver great volumes of information at great speed will change the way managers view what is required for making decisions and controlling their enterprises. It may be necessary, therefore, to experiment before making costly decisions on the type and frequency of information that is required in any given situation.

Although most of this book is concerned with how to structure data in order to produce the information which is required, several of the techniques can be used to help in understanding what information is needed. In particular, the E-R diagram introduced in Chapters 7 and 8 is a highly effective exploratory tool which can be used as the medium for discussion on information needs. In addition, the computer can be effective in clarifying requirements by producing 'rough and ready' systems which can be used by management until their needs become more definite. These prototype systems are cheap, quick and easy to produce using 'fourth generation' software systems. Indeed, computer professionals are increasingly turning to prototype systems as a good method of establishing the information requirements of computer users.

So, what is information? To Stamper (*Computer Journal*) it is a vague generic notion analogous to 'size' (length, weight, volume or mass) in physics. Information has various different meanings depending on how it is measured. At one level it is a measure of information which depends on the relative