Report Series
No. 32

Data Management

November 1982

The Butler Cox Foundation

THE BUTLER COX FOUNDATION

REPORT SERIES NO. 32

DATA MANAGEMENT

ISSUED NOVEMBER 1982

Abstract

Commercial, social and governmental activity in urban societies is becoming increasingly complex and information-dependent, and data provide the raw material for this information. As a consequence, data (not information) are now a corporate resource, and they deserve to be managed deliberately, just like the traditional resources of money, staff, stocks, plant and equipment.

The report defines data management, and sets out its theoretical basis and its impact on systems strategy and end users. It describes the experience to date of managing data, and reviews the available tools and techniques. The report then addresses the management issues of data management, focusing on how the data management function should be organised. Finally, the report advises when data management may or may not be relevant, and sets out guidelines for those organisations wishing to implement a data management policy.

Research team

The team that researched and wrote this report was:

Tony Brewer, a director and partner of Butler Cox, specialising in commercial systems strategy. He was the author of Foundation Report No. 11 — Improving Systems' Productivity, and No. 16 — The Role of the Mainframe in the 1980s.

Rob Moreton, a consultant with Butler Cox specialising in system development and data processing management. He has contributed to the research programmes of several Foundation reports, and was responsible for arranging the programme for two recent Foundation conferences that dealt with systems productivity. He has also lectured extensively on these topics.

Ian Palmer, an independent consultant with an international reputation as a leading authority on database technology and systems development methodology. He served for several years as vice-chairman of the Codasyl Data Description Language Committee, and helped to establish the British Computer Society's Database Administration Working Group.

Photoset and printed in England by Flexiprint Ltd., Worthing, Sussex.

THE BUTLER COX FOUNDATION

Butler Cox & Partners

Butler Cox is an independent management consultancy and research organisation, specialising in the application of information technology within commerce, government and industry. The company offers a wide range of services both to suppliers and users of this technology. The Butler Cox Foundation is a service operated by Butler Cox on behalf of subscribing members.

Objectives of The Foundation

The Butler Cox Foundation sets out to study on behalf of subscribing members the opportunities and possible threats arising from developments in the field of information systems.

New developments in technology offer exciting opportunities — and also pose certain threats — for all organisations, whether in industry, commerce or government. New types of systems, combining computers, telecommunications and automated office equipment, are becoming not only possible, but also economically feasible.

As a result, any manager who is responsible for introducing new systems is confronted with the crucial question of how best to fit these elements together in ways that are effective, practical and economic.

While the equipment is becoming cheaper, the reverse is true of people — and this applies both to the people who design systems and those who make use of them. At the same time, human considerations become even more important as people's attitudes towards their working environment change.

These developments raise new questions for the manager of the information systems function as he seeks to determine and achieve the best economic mix from this technology.

Membership of The Foundation

The majority of organisations participating in the Butler Cox Foundation are large organisations seeking to exploit to the full the most recent developments in information systems technology. An important minority of the membership is formed by suppliers of the technology. The membership is international with participants from the United Kingdom, France, Sweden, Switzerland, Denmark, the Netherlands, Belgium, Italy, South Africa and the United States.

The Foundation research programme

The research programme is planned jointly by Butler Cox and by the member organisations. Each year Butler Cox draws up a short-list of topics that reflects the Foundation's view of the important issues in information systems technology and its application. Member organisations rank the topics according to their own requirements and as a result of this process a mix of topics is determined that the members as a whole wish the research to address.

Before each research project starts there is a further opportunity for members to influence the direction of the research. A detailed description of the project defining its scope and the issues to be addressed is sent to all members for comment.

The report series

The Foundation publishes six reports each year. The reports are intended to be read primarily by senior and middle managers who are concerned with the planning of information systems. They are, however, written in a style that makes them suitable to be read both by line managers and functional managers. The reports concentrate on defining key management issues and on offering advice and guidance on how and when to address those issues.

THE BUTLER COX FOUNDATION

REPORT SERIES NO. 32

DATA MANAGEMENT

CONTENTS

REPORT SYNOPSIS	i
PREFACE	iii
1 THE MEANING OF DATA MANAGEMENT	1
Data and information	1
The importance of data	1
Characteristics of data management	2
Definitions	3
Misconceptions	3
Summary	4
2 THE IMPACT OF DATA MANAGEMENT	5
Theories of data management	5
Impact on systems strategy	7
Impact on end users	9
Summary	9
3 USERS' EXPERIENCE OF DATA MANAGEMENT	10
Case history A	10
Case history B	11
Case history C	12
Case history D	13
Summary	14
4 REVIEW OF TECHNIQUES AND TOOLS	15
Data analysis	15
Data dictionary systems	17
Database management systems	19
Non-procedural languages	22
Data design aids	22
Future developments	22
Summary	23
5 THE ORGANISATION OF DATA MANAGEMENT	24
Policy issues	24
Defining the responsibilities	25
Organising the responsibilities	26
Skills required	27
Data management standards	27
Centralisation and distribution	27
Summary	27

6 JUSTIFYING DATA MANAGEMENT 28
The relevance of data management 28
Costs and benefits of data dictionary systems 28
Costs and benefits of data analysis 29
Costs and benefits of data administration 29
Costs and benefits of database and database management systems 29
Persuading top management 30

7 GUIDELINES FOR ACTION 31
Planning for data management 31
Implementing the plan 31
Human issues 32
Education and training 32
Implementing a data dictionary system 33
Mistakes to avoid 33

CONCLUSION 35

BIBLIOGRAPHY 36

DATA MANAGEMENT

REPORT SYNOPSIS

For historical reasons, mainframe computer systems design originally concentrated on procedure and logic while largely neglecting the effective organisation of data. As a result, the overall picture has been one of proliferating files, complex system architectures, inefficient interfaces, and systems that are unreliable and difficult to maintain.

Today, with the growth of distributed, interactive processing, it is now widely accepted that successful systems depend equally on logical procedures and on accurate and consistent data. Data management has therefore assumed a new importance.

Data management is important because data (the term is used in the plural throughout this report) are a vital corporate resource — just as vital as the traditional resources of capital, labour, stocks, plant and equipment. Data form the basis of the information systems on which organisations depend in order to function effectively, and data management plays a key role in establishing the overall systems framework. Data provide the common currency for this framework and therefore they deserve proper management, with appropriate organisation structures and lines of authority.

To reap the benefits that effective data management makes possible, it is essential that the approach has the wholehearted support and commitment of top management. But first, top management must be convinced that data management is important, that it can be justified, and that it can be introduced successfully.

The introduction of data management can be difficult to justify. Costs will be incurred before any benefit is obtained, and the benefits will be mainly unquantifiable and sometimes intangible. Investments will be needed in manpower, training, education, and tools and techniques. But our research, including case histories (chapter 3), confirmed that the theoretical benefits of data management can be achieved.

Among the specific benefits obtained by the four case-history organisations were:

— Reduced data duplication.

— Increased reliability of data.

— Greater availability of data.

— Reduced time needed for applications development.

— Provision of a foundation for end-user computing.

— Provision of a corporate data model, which provides the basis for integrated systems and for data sharing.

— Involvement of end users in data analysis, thereby enabling them to contribute to data-model and applications development.

— Greater ease and flexibility in database design.

Data management cannot be introduced quickly or easily. A long-term strategy to develop the corporate data resource will be needed, and this strategy may take several years to implement. In moving towards the adoption of data management, a two-stage plan (described in chapter 7) is recommended. First, preliminary work is needed which includes selecting a methodology for system development, defining standards and working methods, applying data analysis to a new system development project, and selecting a database management system (if one is required).

In the second stage, actions include appointing a data administrator, establishing an education programme, applying data analysis at the corporate level and developing a high-level corporate data model, and applying the full system development methodology to all development projects.

As data management is introduced it may incur resistance from technical staff and end users, because changes are required in attitudes and working habits. Resistance may also come from senior managers if they fail to understand the objectives of data management. If this is the case, they will see the changes required by data management as disruptive and they will question its benefits.

Common mistakes that can undermine the introduction of data management include:

—Allocating responsibilities at too low a level.

—Allocating responsibilities which are not clearly defined.

—Failing to co-ordinate the various functional responsibilities, leading to conflicts of interest.

—Failing to establish effective communications between the data management function and the rest of the business.

But perhaps the biggest mistake is the failure to secure determined support from top management. With support from the top, a corporate-wide view of priorities can be taken, and the highly desirable user-involvement can be secured.

Data management will have an impact on organisations' information systems strategies. Additional processing resources will be needed, for example, and there will be changes in the pace and sequence with which applications are developed. There is an impact on end users, also. Increased data management implies increased user involvement, and indeed is an essential prerequisite to end-user computing. These various impacts are discussed in chapter 2 of the report.

The techniques and tools of data management, described in chapter 4, include data analysis, data dictionaries, database management systems, non-procedural languages and data design aids. In database technology, it has been traditional to focus first on the database management system. We recommend a reversal of this approach, concentrating first on data analysis with a data dictionary as a vital tool; secondly on query languages and non-procedural languages as a crucial means of accessing the data; and finally, and only if justified, on a database and database management system.

In many organisations, the concept of managing data as a corporate resource will not be accepted easily. The management and end users need to understand why data are an important resource, and top management then needs to make a firm commitment to promote the introduction of data management. Policies will then be required for allocating the responsibilities for data management, for developing a strategy for the data resource, for defining the ownership of data, for promoting the sharing of data and for defining the scope of data management. These issues are discussed in chapter 5.

But before these issues are even considered, the adoption of data management must be justified. Chapter 6 of the report makes it clear that the case for data management will not be self-evident because it is particularly difficult to quantify the benefits. Data management is likely to benefit most organisations, particularly those that are complex and large, that need flexible management information systems, that have distributed or decentralised computer systems, and possess a proliferation of systems developed by local users, and where there is scope for data to be shared between several users or activities.

There are three basic principles of data management. First, to optimise the sharing of data between different information systems, so simplifying the development of groups of related systems and minimising data redundancy. Second, to maximise the value of data by making them easy to access and manipulate. Third, to achieve both logical data independence (separating the uses of the data from the structure of the data) and physical data independence (separating the method of physical storage from the data structure).

This report discusses how to translate these three principles into practice.

PREFACE

A brief review of the history of data processing clearly shows the extent to which attitudes and concepts have been determined by the available resources. Early systems were constrained by the processing resources then available — single mainframe computers generally located at the centre of the organisation, requiring raw data to be batched up and transported from wherever they originated to the computer for processing. Files of data were stored on sequential storage media, and the cost of equipment was high relative to staff costs. These constraints resulted in an emphasis on computing efficiency rather than on end-user effectiveness. This in turn naturally led to system designers focusing on the problems of procedure and logic, and the early methods for system design and programming reflected this focus. By contrast, there was very little concern with data. Data files were designed to satisfy individual programs and were either recreated from raw data or, at best, sorted into a different sequence to satisfy the needs of other programs.

The attitudes and concepts derived from the design of program-dependent files changed little with the introduction of direct access storage devices, even though these devices allowed program-independent files. As a result, organisations found themselves with a proliferation of files, redundant data and complex system architectures. Moreover, the systems were unreliable and very difficult to maintain and enhance.

Today, the trend is from batch to interactive processing and from centralised to distributed systems. One consequence of this trend is that the traditional emphasis on procedure at the expense of data is now decreasing, as the need for computing efficiency is recognised as being less important than end-user effectiveness. It is now widely recognised that successful systems depend equally on logical procedures and on accurate, consistent data.

Purpose and intended readership

This report shows that the key to accurate, consistent data is data management, and that from a theoretical point of view there is a strong prima facie case for introducing it. From a practical point of view, however, the case is less compelling. For data management to be successful the whole organisation needs to be committed to the concept, and this in turn means that top management needs to be convinced of the benefits it will bring. Three important questions must be answered before the top management in organisations will accept a commitment to data management, with all that it implies. These questions are:

— Have any organisations successfully adopted data management?

— What have been their experiences?

— How can an organisation justify data management?

The purpose of this Foundation report is to provide answers to these questions.

The report is intended for those who will be responsible for devising and implementing a corporate data management strategy. It will also be of interest to senior information systems managers, particularly those responsible for database management systems. The report does not contain details of the tools and techniques of data management. Instead, it provides a management-level perspective of the value and the role of the tools. For this reason, the report should also be of interest to the managers of user departments whose systems are likely to be developed within the framework of a corporate data management strategy.

Structure of the report

The report has seven main chapters. Chapter 1 defines what we mean by data management and, in particular, differentiates between data and information. It also highlights the growing importance of data and information. Chapter 2 examines the potential impact of data management by first reviewing its theoretical basis, and then comparing the theory with what has been achieved in practice. The chapter then identifies the likely impact of data management on systems strategy and on end users.

Chapter 3 summarises our research into users' experience of data management, and presents four representative case histories. The findings of our research into the tools and techniques of data management are set out in chapter 4.

Chapter 5 then turns to the management issues of data management and identifies the lessons to be learnt from the experience to date. The chapter focuses on the ways in which the data management function has been and should be organised.

The report shows clearly that substantial costs must be incurred before any worthwhile benefits can be obtained from data management. For many organisations, the question of how to justify data management will therefore be most significant. This question is addressed in chapter 6.

Finally, in chapter 7, the report gives guidelines for those organisations intending to introduce data management.

A short bibliography concludes the report.

CHAPTER 1

THE MEANING OF DATA MANAGEMENT

In this chapter we define what we mean by data management. We begin by distinguishing between data and information, and then highlight the growing importance of data to all types of organisation. Next, we identify the main characteristics of data management and compare these with the management of other business resources. Because data management is a relatively new discipline, there are several commonly held misconceptions as to what it means and about its terminology. The final sections of the chapter define the terms as used in this report and dispel some of the misconceptions.

DATA AND INFORMATION

Data identify, describe and measure the things (referred to as entities) that exist within or impinge upon their environment. They depend upon linguistic or numeric conventions that have meaning within their environment. They are generally dependent upon coding systems, either implicit (for example, surnames, ages, distances) or explicit (for example, staff numbers, location codes, expense codes).

Data are neutral and have no meaning. They are given meaning by the context in which they are used. If the data and their context are meaningful then they become one form of information. A variety of sensory inputs may also be regarded as other forms of information.

In a commercial environment there are business entities such as customers, orders, products, stocks, invoices, depots, suppliers, etc. They are inherently stable, because they are characteristic of that environment. For groups of similar organisations, such as companies within a particular industry, these entities may even be characteristic of the industry itself. The data that describe these business entities are therefore also stable.

By contrast, the commercial uses of these data are generally much less stable. The uses depend upon external factors like the economic or competitive environment, or on internal factors like company policies, organisation structures, marketing strategies and managerial preferences. In order to make the most effective use of the data it is necessary not only to identify the business entities but also to understand the relationships between them. Clearly defined data and their inter-relationships therefore provide a firm foundation both for management decision making and for systems development.

Data provide the raw material for business information, and different kinds of information are required for different business purposes. At the strategic-planning level many of the problems will not have occurred before and will be unique. They will also be unstructured, so that the strategist will be concerned first with finding a way of solving the problem and only later with obtaining the information to be fed into the proposed solution. The required information will be drawn from many sources, both internal and external, will be aggregated rather than detailed, and availability will be more important than accuracy.

At the operational-control level the problems are more likely to be highly structured, repetitive and therefore well understood. They may be capable of being resolved by automated solutions and decision-making processes. The information required by an operational-level manager will be drawn from a few specific sources, mainly from within his own department; and it will need to be detailed, accurate and up-to-date.

THE IMPORTANCE OF DATA

Commercial, social and governmental activity in urban societies is becoming increasingly complex and information-dependent. In the more advanced industrial nations, more people are now employed in information handling than in primary or secondary industry. Information handling is of particular concern to commercial organisations for three reasons:

— Their competitiveness and flexibility are increasingly dependent upon the quality and availability of information.

— Information derived from data is now a substitute for the kind of direct sensory input that is prevented by size and complexity.

— The cost of managing the data from which that information is derived is a significant proportion of total costs.

Because data provide the raw material for this information, data (not information) are now a corporate resource. They deserve to be managed deliberately, just like the traditional resources of money, staff, stocks, plant and equipment.

CHARACTERISTICS OF DATA MANAGEMENT

The purpose of data management is to ensure that data are available, and are appropriate for the various purposes for which they may be used, at a cost that is justified by the benefits.

As with the management of any other resource, data management entails the following types of activity:

- —Planning, to identify potential uses and hence the supply requirements.
- —Supply , to identify reliable sources.
- —Storage, to provide safe keeping when the resource is not in use.
- —Control, to control the uses to which the resource is put.
- —Security and quality, to prevent leakage or theft, to control access and to ensure adequate quality.
- —Retrieval, to provide easy access for authorised users of the resource.

Figure 1 tabulates the resource management activities for money alongside the equivalent activities for data.

Three basic principles

The first basic principle of data management is to optimise the sharing of data between different information systems. Shared data simplifies the development of groups of related systems, because each system can be designed and installed in an incremental way, with simple interfaces between each one. Data sharing also minimises data redundancy, with consequent reductions in the costs of storage, updating and inconsistency.

The second basic principle is to maximise the value of data, by making them easy to access and easy to manipulate.

The third basic principle is to achieve both logical and physical data independence. These concepts are fundamental to an understanding of data management. Logical independence means making the uses of the data independent of the structure of the data. Thus a programmer's or end user's view of the data structure can be closely related to their own needs and does not have to be distorted by the needs of other applications. New applications may be added without changing the data structure and, conversely, the data structure may be changed without affecting the existing users. These benefits can improve the ease with which applications can be developed and maintained.

Physical data independence means making the method of physical storage independent of the data structure. Improved storage devices can be introduced, or the arrangement of physical storage can be changed to improve efficiency, without affecting the data structure or the existing applications.

Basic principles are easy to state but are difficult to achieve in practice. Although data sharing is no

Figure 1 Resource management activities

Activity	*Money*	*Data*
Planning	Profit/cash planning	Systems strategy
Supply	Internal sales, external loans	Internal sources: — transaction systems External sources: — data banks — informal intelligence
Storage	Bank safe-deposit, loan-out	Databases or files
Control	Accounting and budgeting procedures	Data analysis, data dictionary
Security and quality	Physical restraints, procedural controls	Physical restraints, logical restraints, procedural controls, input vetting, periodic audits
Retrieval	Expenditure authorisations	Report generators, query languages, presentational aids

longer constrained by technical limitations, it is still inhibited by political and inertial factors. Political because the introduction of data management and of data sharing cause shifts in the balance of corporate power. Inertial because systems analysts and programmers — especially those with management responsibility — are uncomfortable with the newer ideas and techniques of data management and find it easier to stick to tried and trusted procedural methods.

Data independence is difficult to achieve because it depends largely on software tools that are still not as powerful as they need to be. Access and manipulation are difficult to achieve mainly because of the way in which existing data are stored.

DEFINITIONS

Certain words and terms are used in a particular way in this report, and to avoid confusion we define them below.

Database: a structured store of data; that is, a store containing both data and the means of maintaining the relationships between the data which reflect the relationships between the real entities described by the data.

Database system: an application system, comprising computerised or clerical procedures and databases which are shared with other application systems.

Database management system: an inter-related set of software tools designed to provide access to a computerised database, and to enforce privacy, security and integrity of the data. The tools also contain special languages for describing the data relationships and for manipulating the data.

Data dictionary system: a special database with its own database management system. Typically, a data dictionary might be used to record:

—The names, descriptions, definitions, attributes and relationships of entities in the organisation.

—The descriptions of procedures and computer programs.

—The descriptions of logical data structures.

—The descriptions of physical data structures.

The data dictionary is therefore an inventory of the organisation's data resource.

Data analysis: the activities of identifying entities and the data that describe them, and representing them in a data model. Data analysis is a useful adjunct to all system development work and is an essential part of data management.

Schema: the word suggested by the Codasyl Database Task Group to describe a model depicting data types and the relationships between them, irrespective of processors or storage devices. Synonyms include conceptual schema, entity model, data model, relationship model, global model and conceptual model.

Sub-schema: the word suggested by the Codasyl Database Task Group to describe a subset of the schema that represents a particular user's view of the data.

Applications framework: an overview of the organisation which highlights the business systems with the greatest potential for computer support and also the entities with the greatest potential for data sharing.

MISCONCEPTIONS

There are many misconceptions associated with data management, and we now describe the three most common ones.

The first misconception is that data management is concerned primarily with using a database, especially a database stored on a computer and controlled by a database management system. For most organisations only a very small proportion of their total data is stored in a database, or even in a computer system. The benefits of data management are gained by analysing, standardising and sharing data. One outcome may be the development and use of a database but this is not an essential characteristic of data management.

The second misconception is that the ultimate objective of data management is to develop a single, immense, computerised database containing all of the organisation's data. One of the basic principles of data management is to promote data sharing, but operational flexibility and efficiency often require data to be duplicated in a controlled way. Also, particular data items may have only tenuous links in the schema or no links at all, so there is no reason for them to be stored in the same database as each other. Data management does not necessarily preclude separate databases, developed to serve particular groups of applications or containing the same data but structured to satisfy different needs, provided that they are all based on a common schema, so that they can subsequently be merged or linked if desired.

The third misconception is that the solution of any management problem can always be assisted by providing more or better information. There are various stages in the use of information which illustrate that lack of data is not necessarily the problem.

To begin with, a manager might say that he does not have the information he needs. The traditional solution was to provide the relevant data along with a mass of irrelevant data in the form of a massive computer printout.

The next stage is for the manager to say that he now has the required information but not the means to manipulate it. The solution to this problem has been to provide more precise access methods (such as menu-driven video terminals) and also data manipulation facilities (such as modelling languages and statistical routines).

Next the manager might say that although he now has the information and the means to manipulate it, the answers do not make sense, because of the poor quality of the data. The solution to this problem is to use data management to ensure that the data are available, consistent and with the required levels of accuracy.

But the manager may then finally realise that he does not know how to solve the problem. The difficulty here is that certain management problems, particularly at the strategic-planning level, are unstructured and occur infrequently. Solving such problems depends much more on understanding the problem and formulating an appropriate model than on processing existing information.

The conclusion is that, although more or better information frequently leads to better decision making, it is not necessarily the only ingredient. Access to external data and improving managers' abilities to analyse decisions may also be necessary.

SUMMARY

Data management is the formal collection, storage and preparation of data to ensure that they are available and appropriate for the uses to which they may be put. It is important because effective information systems are essential in today's complex organisations, and these systems cannot be developed without careful consideration of the data upon which they depend.

Data management uses techniques such as data analysis, and software tools such as data dictionaries and database management systems. It frequently, but not necessarily, leads to the development of computerised databases.

Successful data management requires a recognition at the highest level in the organisation that data are an important corporate resource. This implies that the data deserve proper management, with appropriate organisation structures and lines of authority. Data management therefore requires the development of a long-term strategy for the development of the corporate data resource. This strategy will probably take several years to implement. It will involve the introduction and use of data analysis, the setting up of management units responsible for the control and use of data, the building of data models, the selection of software tools and the formulation of an implementation plan for creating and consolidating databases.

CHAPTER 2

THE IMPACT OF DATA MANAGEMENT

THEORIES OF DATA MANAGEMENT

Two useful theories, describing and explaining the development and impact of data management, have been suggested and are summarised here.

Nolan's theory

In his article "Managing the crises in data processing" (Harvard Business Review, March-April 1979) Richard L. Nolan formulated his six-stage theory of data processing development. People may question his assertion that all organisations using data processing must evolve through all of the stages that he postulated, but his theory does provide a model that is easy to understand and which is a very useful reference.

In summary, he suggested that the use of data processing by an organisation evolves through six stages:

—Initiation.

—Contagion.

—Control.

—Integration.

—Data administration.

—Maturity.

In the first three stages the emphasis is on the management of computing resources, but it then changes to the management of corporate data resources. Each stage is characterised by the relative emphasis of management attitude between "slack" and "control". A slack attitude encourages innovation and initiative but leads to rapid and uncontrolled growth. An attitude of control encourages effectiveness and efficiency but leads to the containment or balancing of supply and demand. The management attitude towards computing resources changes during stages one to three from slack to control, and a similar change in attitude towards data resources occurs during stages four to six.

Stage three is the transition stage. Initially there is an emphasis on tight control of computing resources, leading to the development of internal standards, an increase in planning and project management and the introduction of charges for computing services.

However, it is precisely this analysis of computing as a resource to be managed that leads to the conclusion that an overall systems framework is needed. Such a framework is necessary if individual systems are to be integrated effectively and the needs of management for control and planning information are to be met. It becomes clear that data provide the common currency for this systems framework and therefore that the data must be actively managed.

Stage four is characterised by rapid growth in a slack environment, but leads to data duplication and inconsistency. In stage five the balance of management attitude swings from slack towards control, with the introduction of data administration. Finally, in stage six — maturity — there is an equilibrium between resource supply and demand, with data highly managed but also freely available to authorised users.

From the point of view of Nolan's theory, data management has the following impacts:

—A change from the planning and control of computing resources, exercised within the data processing department, to the planning and control of data resources, exercised outside the department by users (through recharge mechanisms), by data administration and by a top management steering group.

—A balance established between central control both of data and common application systems (generally at the operational-control level), and end-user control of data retrieval and manipulation systems (generally at the management-control and strategic-planning levels).

—An evolution of end-user involvement, which tends to be reactive in stages one and two, dominant in stages three and four and participatory in stages five and six.

Palmer's theory

At the Computing Workshop conference held in London in 1980 Ian Palmer described six levels of achievement in data management. Unlike Nolan, he was not implying any process of evolution between the levels (although many organisations do evolve from one level to the next). His six levels are characterised most clearly by the use of databases and database management systems. They are:

First level: file databases
A conventional file has been converted to an equivalent database and a database management system is used to provide access. There is no data sharing except between file maintenance programs and simple enquiry programs. Data analysis is used only to document existing data fields.

Most of the implementation and running costs of a database management system are incurred, and the benefits are likely to be improvements in access to the data, programmer productivity, and data security.

Second level: application databases
A database is created to support a specific application, with little consideration of data independence or long-term flexibility. Data are shared within the application but are not readily available to other applications. Data analysis is limited to normalisation of records from conventional files prior to database design. (The meaning of the term 'normalisation' is explained on page 15.)

In addition to the costs of implementing the database management system there are the costs of performance monitoring, database tuning and the planning of data backup and recovery. The benefits are likely to be the support of complex relationships (bill of materials, for example) and the ease of developing interactive enquiries.

Third level: subject databases
A database is created to support a particular subject (for example, customers, suppliers, or products) with the emphasis on the attributes and relationships of the data within the business, rather than on its use. Data are shared between departments and between applications. Consideration is given to the desirability of logical and physical data independence, and the need for a database administrator is recognised. Data analysis is used to identify data types and their attributes and relationships.

The problems associated with shared data have to be resolved. The cost of increased data security is incurred. The benefits are faster application development and reduced data redundancy.

Fourth level: integrated databases
Subject databases are combined into integrated databases at the logical level, and possibly at the physical level also, based upon a corporate data model. Data are shared between related applications with a minimum of duplication. Data analysis is essential in developing the corporate data model.

Additional costs are incurred in correcting inaccurate or inconsistent data from conventional files and in redeveloping existing systems. The benefits are improved data quality, easier access and greater flexibility to satisfy changes either in requirements or in organisational structure.

Fifth level: corporate databases
Data are regarded as a single shared resource, organised for the benefit of the organisation as a whole. Integration of databases is extended and the databases are shared between all applications, with duplication allowed only to satisfy performance requirements.

Data administration is essential to resolve political issues, and data analysis (supported by a data dictionary system) is essential to standardise and document the data resource.

Additional organisational and managerial costs are incurred but operational costs and system development costs are reduced.

Sixth level: distributed databases
The data resource is physically distributed to wherever it is most frequently used, but logical control is maintained centrally. Data sharing is achieved through data management software, using a communications network.

Data analysis now becomes concerned with the geographical location of data types and procedures.

Additional costs are incurred in terms of software and communications facilities. The benefits are improved performance and the introduction of local responsibility for data maintenance and retrieval.

Some additional characteristics of the six levels are set out in figure 2.

Progress measured against the two theories

Most organisations, if they are using databases and database management systems at all, are at Palmer's levels one or two. This is equivalent to Nolan's stages three and four, and reflects a technical orientation towards data management reinforced by pressure from end users to produce results quickly. This level of achievement may even have been achieved almost by accident, in that growing technical interest in database management systems has coincided with the need to develop new systems.

Progress to the higher levels has, to some extent, been constrained by the particular database

Figure 2 Palmer's six levels of database achievement

Characteristic	*Level*					
	1	*2*	*3*	*4*	*5*	*6*
Data sharing	Simple on-line usage	On-line and batch users	User departments relevant to the subject	Related applications	All applications and users	Local and central users
Data structures	Multiple indexes and several databases	Limited number of records and linkages	Several databases	Linked hierarchies	Complex network	Local networks within central control
Data analysis	Existing data elements	Normalisation of conventional files	Definitions and properties of data elements	Data modelling and functional analysis	Documentation of the data resource	Location of entities
Use of data dictionary	Fields in existing files	Database structure	Database, transactions and usage	As an on-line analysis and design tool	As a complete information system	Location of data and functions
First involvement of . . .	Application programmers, systems programmers	System designers	System analysts	End users	Top management	Data communication technicians
Responsibility for database	Project manager	Systems programmer	Technical database administrator	Political database administrator	Data administrator	Central and local data administrators
Optimisation of . . .	Ease of access to data	Performance for one application	Availability of information on a specific subject	Flexibility by modelling the business	Overall control and consistency of data	Transfer of data
Potential benefits	Less program maintenance, improved security	Support for complex relationships	Faster response to user needs; limited MIS	Flexibility to support changes in the business; control of redundancy	Decision support system for top management	Local responsibility for data quality and usage
Likely problems	Back-up and recovery strategy	Performance monitoring; database reorganisation	Unclear responsibilities; data security; test strategies	New methodologies, redesign and reprogramming; inconsistent data	Ownership of data; organisational politics; database restructuring	Local independence; migration of data

management system used. IMS users have often moved quickly to level three, for example, but have not been able to progress any further. This is because IMS is well-suited to the development of subject databases, which can be expressed as hierarchies, but it does not easily support the network structures needed to represent integrated databases. By contrast, some organisations using database management systems based on Codasyl or network structures (such as IDMS or Total), have progressed to level four.

We know of no organisations that have yet reached Palmer's level five — equivalent to Nolan's level six — although several are planning to do so. The amount of data analysis, data design and data collection required, with the need for consistent policies for data management, mean that most organisations will take many years to reach this level.

Level six is not yet possible, because adequate software giving full physical and geographical independence is not yet available.

We would stress, however, that progress with data management has been limited more by lack of understanding and commitment by top management than by technical factors.

IMPACT ON SYSTEMS STRATEGY

Many organisations have formulated a long-term, strategic plan for developing their information systems to satisfy their business needs. Such a plan generally takes into account developments both in data processing technology (such as minicomputers, distributed processing and packaged software) and in communications technology (such as public data services and computerised branch exchanges). Sometimes the plan also takes account of developments in office systems, such as electronic mail and text processing. Because data management is a relatively new concept, not many organisations will have considered it when developing a systems strategy. Nevertheless, we believe that this situation will change, and we discuss below the impact that data management is likely to have on the main components of a systems strategy.

Applications framework

In chapter 1 we defined the applications framework as an overview which highlights the business systems with the greatest potential for computer support, and also the entities with the greatest potential for data sharing. The overview is depicted by means of:

—A high-level data model, showing the major data types and the relationships between them.

—A function hierarchy, showing the major business systems, broken down to about three levels of detail.

—An entity/function matrix, showing which entities are used in which business systems.

—A data flow diagram, showing how data flows between functions and hence the dependencies between the various functions.

The applications framework is one of the keys to successful data management because it provides a high-level model which links data with their uses, and is independent of any considerations of equipment or software. It also provides a means of ensuring that individual applications can (and do) share data and that databases are integrated at least at the logical level.

Applications portfolio

In theory, data management should have no impact on the portfolio of potential applications, since this is dictated by the needs of the business. In practice, high-level data analysis may diagnose the reasons behind some long-felt irritation, such as inconsistent data or contradictory data definitions and, as a result, additional development projects will be set up to cure the problem.

Processing resources

Additional processing resources will be required to drive the data dictionary system and the database management system and to store an increasing proportion of the corporate data. In theory, this increase should be offset by storage savings created by reducing data duplication, and by processing savings arising from better-integrated applications.

System development and maintenance resources

Initially, additional resources will be required to carry out data analysis and database administration, while existing staff are being trained in the new methods. Additional software tools, such as a data dictionary system, a database management system and data design aids may also be required.

In theory, the number of systems development and maintenance staff should eventually be greatly reduced as productivity rises, development becomes quicker and easier, the maintenance load reduces and end users satisfy many of their own requirements. Whether this will happen in practice remains open to question.

Management control

Data management should lead to much greater management control over the systems strategy, especially in a distributed environment. It will increase freedom of the uses to which data may be put, but will increase control over the meaning of data, their storage, access and security.

Given the software tools available today, data management is inevitably a centralising influence although in theory it should be neutral.

Sequence and pace of implementation

The sequence in which applications are developed is determined partly by the needs of the business and partly by the applications framework. This framework is one of the keys to successful data management, and it may well change the sequence in which applications are developed.

Organisations introducing data management ideally should carry out some high-level data analysis. This analysis should be independent of specific applications, and should be aimed at developing at least the outline of a corporate data model. Subsequently, detailed data analysis must be carried out prior to file or database design for specific application development projects. Because of this two-stage approach to data analysis, the pace of implementation for application projects will be slowed, at least initially. Subsequent development, however, should be much faster.

Justification

In some senses, data management will make it harder to justify the required investment in information systems. In other senses, it will make it easier.

If information systems are assessed purely in terms of quantified financial costs and benefits, then the investment required by data management increases the costs and makes the benefits more difficult to quantify, and so justification will be much harder. However, very few organisations take such a rigid approach to justifying their investments. (Foundation Report No. 24 — Investment in Systems — contains a full description of the methods actually used to justify investments.)

On the other hand, if top management understands the concepts of data management then justifying investments in systems should be much easier. If data management is practised, top management's perception of data processing changes from a low-level productivity issue to a concern with corporate resource management at the highest possible level.

IMPACT ON END USERS

We have already suggested, in describing Nolan's six-stage theory, that one of the characteristics of data management is increased user involvement in information systems. This increase arises for three reasons:

—Data analysis is a simple-to-learn, easy-to-use, technology-free activity. It satisfies the need of end users to be involved in the development of their own systems. It is also a very powerful tool for uncovering long-standing logical contradictions within the business.

—After the initial period of building up the data management infrastructure, system development is quicker and easier, and the need for system maintenance is reduced.

—Data management is an essential prerequisite to end-user computing using either interactive terminals or personal microcomputers. It enables control to be achieved over the use of corporate data, but it also increases the ease with which end users can access the data that they need.

One political disadvantage, from an end user's point of view, is that the lines of communication may become less clear. The end user may be confused by the need to talk either to local or central systems analysts, or to a data administrator or to a database administrator.

SUMMARY

In this chapter we have described the impacts that data management should have in an organisation. The theories described, and the impacts both on the systems strategy and on end users, represent the ideal situation. But what has been achieved in practice? In the next chapter we set out four case histories which we believe are representative of data management as it is practised today.

CHAPTER 3

USERS' EXPERIENCE OF DATA MANAGEMENT

One of the objectives of the research for this report was to find out about users' experience of data management. We followed three lines of investigation.

First, we carried out a survey using a postal questionnaire. A very detailed set of questions was sent to about 300 large organisations, including all Foundation members. We received 50 replies in time to be included in our analysis, and relevant results are included in this report at appropriate places. The profile of respondents is shown in figure 3.

Second, we conducted interviews with a representative sample of organisations who we knew had either already adopted data management or were planning to do so. From these interviews we have prepared four case histories. These are anonymous in order to preserve confidentiality, but the facts are all as stated to us.

Third, we conducted interviews with the major suppliers of software tools in the United Kingdom. The results of this part of our survey are set out in chapter 4.

CASE HISTORY A

Company A is a publicly owned regional utility supplying energy services to some 2.2 million customers in one geographical region of the United Kingdom. It has a large data processing department employing about 320 staff (including 70 systems development staff), and using two large IBM-compatible computers which support over 1,000 interactive terminals.

In the late 1970s the management board formulated an information systems strategy which recognised the importance of data and the need to exploit information technology. A planning exercise was carried out (using IBM's Business Systems Planning methodology), which led to the creation of a data management section in 1979.

The data management section comprises five staff and is headed by the database manager, who reports with four other senior managers to the manager of computer services. The main activities of the section are to promote the use of data analysis, to design databases and to maintain the data dictionary.

Tools and techniques

Soon after the data management section was established, a data analysis exercise was carried out to prepare a data model for the whole organisation. Preparation of the model, which was based on the existing organisation structure, involved interviewing key people and identifying the data associated with their jobs. The resulting data model is used as the basis for individual application development projects.

At the start of each project the data model is examined and the relevant parts are extracted. Each main end user of the application then spends a period working full time with a systems analyst. This period starts with a half-day familiarisation with the principles of data analysis and ends with the production of a detailed data model, which is then coded and stored in the data dictionary. The systems analyst prepares a set of access profiles for the proposed system and a database designer finally designs either a physical database or a conventional file.

The data dictionary system used is Data Manager. This product was selected before the database management system was chosen because it could be used independently of a database management system but has interfaces to the more common systems. One drawback of Data Manager is that its output is in tabular format, but a complementary output-generator product called Diagrafs is used to produce graphical output. Diagrafs is used to print data models in a diagrammatic form that is readily understood both by technical staff and by end users.

The database management system used is IMS/DB. This gives great flexibility and, if correctly used, the run-time performance is as good as that from conventional virtual storage access method (VSAM) files.

Main data areas and applications

The systems that have been developed encompass the main data types of appliances, parts, customer addresses and employees. There are 717 entries in the data dictionary.

Figure 3 Profile of the 50 questionnaire respondents

Annual sales (or expenditure)

$M	*Respondents*
0 - 99	6
100 - 999	15
1,000 - 1,999	9
2,000 +	11
Not available	9

Staff employed

Number	*Respondents*
0 - 4,999	17
5,000 - 9,999	8
10,000 - 49,999	20
50,000 +	3
Not available	2

Data processing expenditure

1982 budget, $M	*Respondents*
0 - 1.9	1
2 - 9.9	28
10 - 19.9	7
20 +	6
Not available	8

Ratio of data processing expenditure to annual sales (1982)

Ratio	*Respondents*
less than 0.5	9
0.5 - 0.99	12
1.0 - 1.49	5
1.5 - 1.99	3
2.0 +	5
Not available	16

Ratio of data processing staff costs to total data processing expenditure (1982)

Ratio	*Respondents*
0.21 - 0.30	4
0.31 - 0.40	16
0.41 - 0.50	7
0.51 - 0.60	13
0.61 - 0.70	1
Not available	9

Computing equipment used

Supplier	*Respondents*
IBM (including IBM compatible)	31
ICL	11
Univac	2
Tandem	2
Honeywell	2
Burroughs	2
Hewlett Packard	2
Others	6

Database management system used

Product	*Respondents*
IDMS (ICL)	9
Adabas (Software AG)	7
IMS (IBM)	10
DL/1 (IBM)	5
Total (Cincom)	4
IDMS (Cullinane)	5
Others	18
No product	7

Data dictionary system used

Product	*Respondents*
Data Manager (MSP)	9
DDS (ICL)	7
IDD (Cullinane)	5
Data Dictionary (IBM)	4
Others	7
No product	18

Justification

The company believes that the main benefits that have resulted from the adoption of data management are:

— User interest is stimulated and sustained.

— Users gain a better understanding of the corporate data and so can state their requirements more clearly.

— Data sharing is promoted, both within the regional organisation and between it and other regions.

— The database design tasks can be performed more easily and with greater flexibility.

Data management has been introduced into this company as a matter of policy. A considerable investment has been made in organisation, in training and in the global data model, but the benefits are now being gained. The main criticism expressed to us was that the company is not proceeding fast enough in implementing its data management strategy.

CASE HISTORY B

Company B is a major engineering company in the aircraft industry. It is divided into nine divisions each with a management board. There are about 500 system development staff in total, and about 130 of these are located in a strong central systems department, which is responsible for developing common systems to be used in all the divisions.

During the last three years a major review of computing has been carried out and the decision has been taken to redevelop the existing systems. These had evolved in a somewhat piecemeal way and had led to the creation of many independent databases with much duplicated data.

The intention now is to develop an overall business architecture that will provide a framework for systems development over the next 20 years. The architecture will encompass all systems at all locations, and individual application architectures will be created within the overall architecture.

Data management activities are the responsibility of a special projects team which is part of the central systems department. This team has ten staff and is primarily responsible for database and teleprocessing support.

Tools and techniques

The term "data analysis" is avoided, since it is felt to have too strong an association with databases.

Nevertheless, data flow diagrams and entity data models form the basis of the special project team's analytical work. Business analysts, working with board-level executive assistants, first define the requirements of the business systems and produce a logical system design using data flow diagrams and entity data models. Technical design then takes place to develop the entity models and, finally, conventional files or physical databases are designed.

The team evaluated IBM's data dictionary system, but rejected it because it was not sufficiently user-friendly and did not support their interactive requirements. The team members subsequently developed their own data dictionary system. The database management system used is Cullinane's IDMS.

Main data areas and applications

The main applications are concerned with production management, including order processing, engineering assemblies and production control. A typical factory system has 50 interactive terminals handling 3,000 transactions per hour. A typical database has 100,000 records and would occupy 200M bytes of disc storage.

Justification

Data has been recognised as a corporate resource in this company, and databases have been developed because they provide the only means of satisfying the requirements for data storage, flexibility and integrity. Data management is now being introduced for the somewhat negative reason that the piecemeal development of database systems has led to data duplication. In addition, the existing database systems have not promoted data sharing and do not provide an adequate basis for the development of systems in the future.

CASE HISTORY C

Company C is an international airline. It uses more than 8,000 interactive terminals worldwide which generate over 1,000,000 transactions per day. Its whole business depends upon information systems and the data which drive these systems are recognised as a corporate resource.

Databases have been used increasingly since 1974 and the various systems share much common data. One important consideration is that the databases should be designed to be independent of the company's organisation structure and working methods, because these are frequently changed. In this sense, the data must be logically independent.

Several different makes of computer are used within this airline and the storage requirements of any particular computer must not be allowed to obscure the underlying structure of the data, which is determined by the business itself. In this sense, the data must also be physically independent.

The company is planning to redevelop several of its systems over the next few years and to integrate its databases.

For all these reasons, the company decided to introduce data management and, in particular, to develop a corporate data model.

Tools and techniques

The database management systems used are IMS on IBM computers, DMS on Univac computers and Enscribe on Tandem computers.

The corporate data model was completed in early 1979. The company places great stress on the contributions both of systems analysts and of end users towards developing application models based on the corporate data model. The six steps required to develop a data model are as follows:

1. Initial seminar
The initial seminar begins with a five-hour tutorial for users and systems analysts. Its purpose is to introduce the concepts of databases, data analysis and model building and to allow the participants to build a simple model from a system description. This is followed by a one-and-a-half day session. After a short briefing about the system for which the data model is being developed, the seminar divides into two groups, each working separately but on the same tasks, and meeting every two hours to check on progress. The groups list the objectives of the system, prepare a first draft of the data model and check it against the objectives. Finally, the seminar reassembles to summarise the results and agree on the next steps.

2. First draft
Immediately after the initial seminar, a fair copy of the first draft of the data model is drawn and distributed to those who attended the seminar. For very large systems the model is broken down into several smaller models, with each one properly related to the main model.

3. Detailed analysis
The first draft of the data model is then analysed in detail by a database designer and by the systems analysts concerned with the system. New data items or relationships are identified and added to the model. This process may take several weeks.

4. Second seminar
The users and analysts reconvene for a second

seminar at which the more detailed draft of the data model is checked and agreed. The output from this stage is a logical data model.

5. *Physical design*
After the logical data model has been agreed, the physical model is designed, taking account of the computer and the database management system that will be used to implement it. Access times and other performance requirements are also taken into account at this stage.

6. *Checking the physical model against the logical model*
Finally, the physical model is checked against the logical model and every difference is assessed. This final check ensures that discrepancies are due to performance requirements or limitations of the database management system rather than to misunderstandings or faults in the model.

Main data areas and applications

The main data types are aircraft, passengers, cargoes, take-off/landing stations, workshops, materials, staff and capital. The applications cover all aspects of the operations of an international airline.

Justification

This airline has identified the following benefits of adopting a data management approach:

- —The company understands its data and is able to manage them as an independent resource.
- —The data analysis method involves end users and enables them to make a real contribution to designing data models and developing applications.
- —The corporate data model provides the basis for integrated systems and for data sharing.
- —The corporate databases provide the performance and flexibility required by the business.

CASE HISTORY D

Company D is the systems company of a major car manufacturing group. Its principal services are consultancy and the provision of a comprehensive processing and network facility, using IBM equipment. There are about 500 systems development staff who are located centrally and at major operating sites.

During the past 12 to 18 months a major review of data, information and database systems has been undertaken. As a result, a data strategy has been defined for the operating companies within the group. The strategy has been accepted at senior management level and its implementation is under way.

The objectives of the strategy are to define the responsibilities for data management within the group, and to establish a structure and some of the procedures for carrying out those responsibilities. The emphasis of the strategy is on the need to manage and administer all data, irrespective of whether or not they are held on a computer system.

The company proposes to appoint data administrators who will each be responsible for a designated 'operational unit'. An operational unit may be a collection of people, departments, functions or companies. The boundaries of the unit determine the level of responsibility for the management of data. The appointments will be at a senior level, at least equivalent to that of an internal company auditor.

Tools and techniques

The data created and used by an operational unit are managed by the data administrator responsible for that unit. The priorities and loyalties of the data administrator therefore reflect the business needs of the unit. The data administrator's primary role is to establish a favourable environment within his own unit that will promote a rigorous and sharing attitude to data, whilst at the same time allowing greater co-operation between operational units.

This approach to data management was developed as the result of several years of large-scale database usage. The company has developed its own data dictionary system, and the database management systems used are IMS and Total.

To perform his role, the data administrator needs to establish and maintain contact with:

- —Functional managers within the operational unit.
- —Data administrators responsible for other operational units.
- —Appropriate systems development staff and database administration staff.

The detailed responsibilities of the data administrator lie in the four areas of documentation, control, support and communications.

Documentation
The data administrator uses a data dictionary system to document the data and the rules for using and accessing them.

Control
The data administrator ensures that all changes to the data characteristics (but not values) are notified by the person making the change. In addition, he establishes who is the prime author (owner) of the data, and he defines the associated privacy and security requirements.

Support
Because he is the operational unit's data expert, the data administrator can play a significant role in identifying new applications which may have a direct bearing on the unit's efficiency and profitability. He will also help systems staff with data analysis, and contribute to the preparation of logical data structures. In essence, the data administrator co-ordinates the interaction between users and systems staff as they set about establishing integrated databases and systems.

Communications
The data administrator identifies user-education needs and ensures that these needs are met. He also meets frequently with other data administrators to ensure that a unified approach to data management is being taken throughout the whole group.

Main data areas and applications

More than 40 application areas have been implemented on databases including distributor support, sales order processing, vehicle specification and parts control.

The plan is for all operating companies within the group to adapt their organisation in line with the data management strategy as quickly as local conditions allow. When this report was written, two data administrators had been appointed.

Justification

The company realised that significant costs would have to be incurred in order to establish a data administration function. These costs are not only for staff but also for software and computer resources. A pilot project was therefore undertaken in order to gain experience of performing the task of data administration, and to monitor the costs and benefits.

The company also recognises that, although the strategy is in the long-term interests of the operating companies, economic pressures may make full implementation difficult in the short term.

The benefits of adopting a data management strategy may be summarised as follows:

- —A solid basis for developing and operating information systems has been provided.
- —Data that are required for mandatory legal and audit requirements, together with data required for decision making, will be more readily available.
- —The amount of duplicated (and often conflicting) data will be reduced, together with the effort required to maintain the data.
- —Data should be more reliable, because they are managed through a function which has no sectional bias.
- —Systems staff will spend less time on data analysis, leading to shorter timescales for implementing new applications.
- —A foundation of well-managed data has been provided for the development of end-user computing.

SUMMARY

The following five points stand out clearly from these four case histories:

- —Data management had not been introduced by accident or by default. In all cases it had been introduced either as an aspect of strategic systems planning or as a solution to a severe difficulty or in response to a need within the organisation.
- —In all cases data management had top-level support.
- —In all cases end users were closely involved in data analysis.
- —In all cases data are recognised as a corporate resource.
- —In three of the cases the benefits, although not quantified, confirmed the validity of the original decision.

Our interviews with user organisations confirmed that the theoretical benefits of data management can be achieved. In the next chapter we review the tools and techniques that are being used to achieve those benefits.

CHAPTER 4

REVIEW OF TECHNIQUES AND TOOLS

At the beginning of this report we argued that early system design and programming methods over-emphasised procedure and logic and gave insufficient attention to data. We suggested that the reasons for this attitude are now decreasing, with the trend from batch to interactive processing and from centralised to distributed systems. At the same time, the emphasis on the quality of data as the basis for reliable information is increasing, and so the need to manage data has been recognised. In this chapter we review the techniques and tools of data management. We discuss them under the headings of data analysis, data dictionary systems, database management systems, non-procedural languages and data design aids. We then predict likely future developments in data management techniques and tools.

DATA ANALYSIS

In the analysis stage of the system life cycle it is necessary to have a methodology that looks at the organisation both in terms of the procedures carried out (the traditional concern of systems analysis) and in terms of the data required. This second activity is known as data analysis, which in turn divides into the two main activities described below.

Identifying entities and their inter-relationships

The first activity in data analysis is to examine the organisation to identify the principal entities upon which it depends — suppliers, customers, products, branches, orders, invoices, accounts, etc. Then, unlike procedure analysis which asks what the entities do or what happens to them, it asks what they are, what are their attributes and their relationships with each other. Finally, it looks for the data which describe these entities and their attributes and relationships.

This activity is often referred to as entity analysis. We stress that it is not a conceptual, desk-based activity, dealing with data at one level of abstraction from the real world. It is an activity which is absolutely concerned with the real world and which end users can and should do for themselves. It frequently uncovers different entities with the same name, or the same entity having many different meanings. For example, an oil company discovered that the entity "port" had many different meanings in different parts of the company; a government department discovered that many different human relationships were referred to in the legislation by the single entity "married couple"; and a retail company discovered that confusion about the meaning of the attribute "purchase price" had always undermined the validity of its management accounts. These confusions had never been uncovered by traditional organisation and methods (O&M) analysis nor by procedure analysis.

One of the important formal techniques within this part of data analysis is known as normalisation. This involves breaking down complex data structures into simple structures in flat-file format. It is a teasing-out process which reveals the true attributes and relationships of the data. Fully normalised data are said to be in third normal form (although some experts argue that a fourth normal form is necessary).

The data model

The second main activity in data analysis is to depict the entities and their inter-relationships in diagrammatic form in a data model. The data model provides the basis for subsequent data design, and its creation may well reveal omissions, duplications and contradictions. A variety of modelling techniques is available. Some are important from a conceptual point of view but are of little practical use; others have been in practical use for many years.

The better known modelling techniques are listed in figure 4. Techniques derived from the Bachman role

Figure 4 Data modelling techniques

Technique	*Basis*
Bachman	Role model
Chen	Entity relationship model
Codd	Relational model
Finkelstein & Martin	Information engineering
Hammer & Mcleod	Semantic data model
Lundeberg	Information analysis and activity analysis
Palmer	Entity analysis and function analysis
Smith & Smith	Aggregation and generalisation

model and the Chen entity relationship model are probably the most widely used. The relational model postulated by Ted Codd depicts entities as relations in two-dimensional tabular format. It has the advantage of being conceptually simple and mathematically precise, but is not well suited to business analysis. Nevertheless, if data analysis has been carried out using other more powerful techniques, the results can be mapped onto the relational model as a preparation for using a relational database management system (see also page 20).

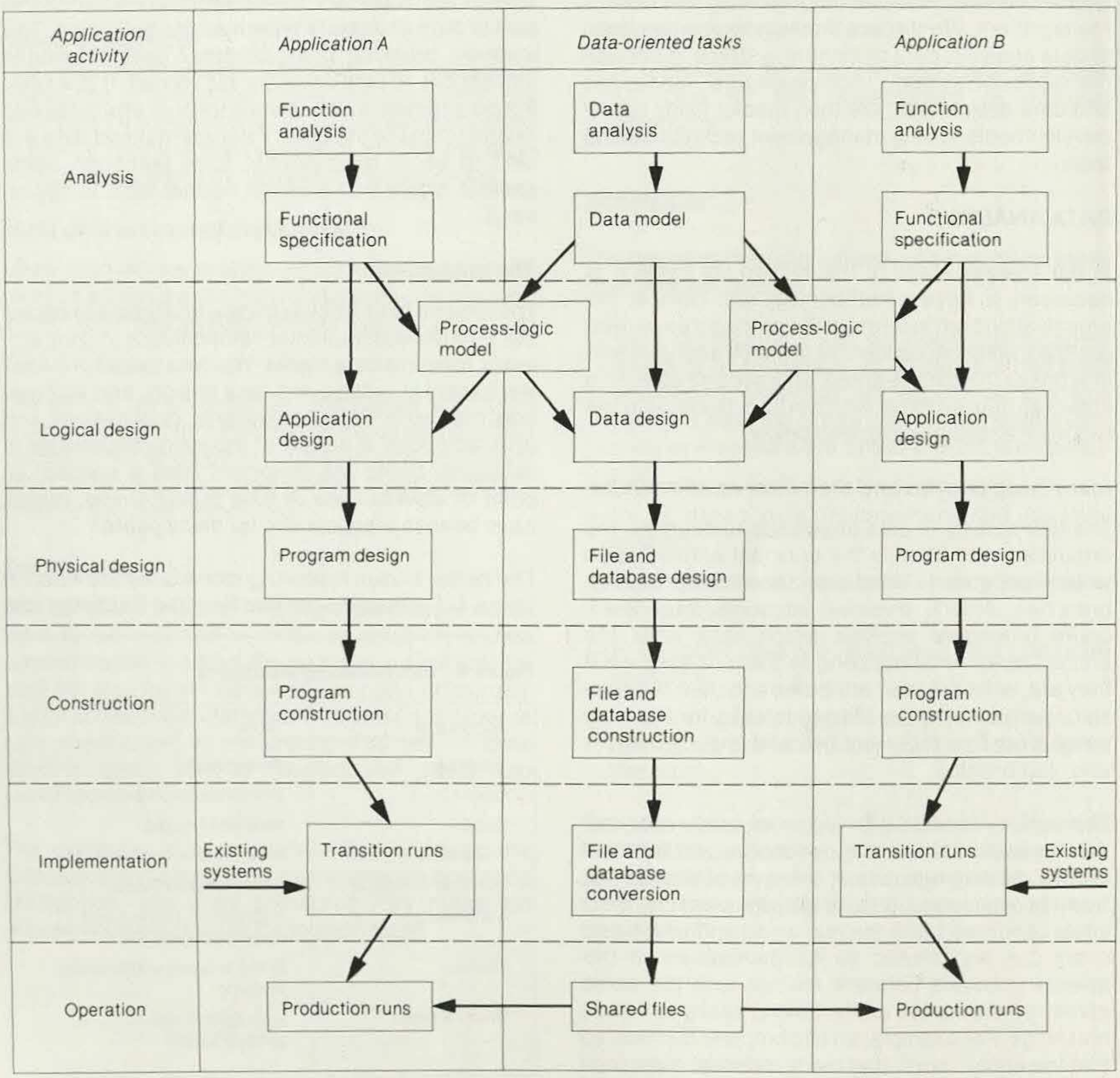
Applying data analysis

Data analysis is essential to data management because it locates, defines and standardises the data resource in preparation for data sharing between applications.

Figure 5 illustrates the way in which data analysis can support the development of two separate applications. During the analysis stage, function analysis in each of the application areas leads to a functional specification of requirements. Data analysis proceeds in parallel, using the global data model as a high-level reference, leading to a data model for each application area.

During the logical-design stage a process-logic model is created, for each application, which combines the data model and the functional specifi-

Figure 5 Data analysis and design support for application development



cation. This then leads to the logical design of the database or files (using the global data model as a reference) and the logical design of the application (using the application framework as a reference).

During the physical-design stage the requirements of processing, equipment and software are taken into account, and physical designs are prepared for the database or files and the programs. This stage is followed by the stages of construction, implementation and operation.

The logical-design stage of data analysis provides the opportunity for sharing data between the applications and (possibly) for integrating the systems also.

As our case histories demonstrate, users have found that data analysis is not only an essential preliminary to good data design, but it can also help to focus attention on business problems. In addition, it has proved to be an excellent way of involving end users in the development of their systems and of allaying their fears about losing control of their data.

There are some pitfalls to be avoided in the use of data analysis, however. These include:

- —Restricting the scope of the analysis to a single application area.
- —Concentrating too much on existing data, especially those contained in computerised files, and neglecting real-life business entities.
- —Concentrating on the attributes that are of interest to only one application, and so not shareable.
- —Confusing analysis of the business with logical design of the data. (This pitfall is encouraged if the data dictionary system does not provide support at both levels.)
- —Regarding normalisation as the solution to all data analysis problems. Normalisation can only be as useful as the analyst's understanding of the entities or attributes being normalised.
- —Failing to change data models to reflect changes in the business.

User experience of data analysis

In our survey, we found that nearly two-thirds of the organisations responding claimed to be using data analysis, and they gave data analysis a satisfaction rating of 3.7 (on a scale of 0 to 5). Amongst these organisations, data analysis had been used for an average of more than three years, and it was a mandatory technique for 67 per cent of them. Nearly 90 per cent of the organisations using data analysis had used external consultancy support. This evidence suggests that data analysis is a widely used and successful technique. But the same users also said that, on average, seven staff were experienced with data analysis, and that only four applications and three databases had been developed using the technique. Moreover, only 36 per cent of them had updated their standards to include data analysis. This contrary evidence tends to support that obtained from an EEC survey in 1979, which suggested that data analysis was not, in fact, widely used. Perhaps the truth is that many organisations believe in the importance of data analysis in principle but fail to make extensive use of it in practice.

DATA DICTIONARY SYSTEMS

The responsibilities of management cannot be carried out without the necessary information, and data management is no exception. The information used by the data administrator and database administrator is derived from data which describe the corporate resource 'data'. These data about data are known as metadata.

Metadata come in three varieties. Semantic metadata describe the meaning of data. They include identifiers, definitions, synonyms and descriptions. Physical metadata describe the means of representing the data. They include the field size, field type, storage medium, frequency of occurrence and frequency of use. Usage metadata describe what the data are used for and by whom.

Metadata must be collected and recorded in a way that enables them to be easily maintained and used. A data dictionary system is designed for this task and is a vital tool for data management. It manages the metadata used within data processing and particularly by the data administrator, whereas a database management system manages the data used throughout the organisation.

The term "data dictionary system" is a poor description of the contents and purpose of this type of system, because it implies merely a list of data types, each with its definition. This is very far from the whole truth. A data dictionary system can be used to document metadata at the business level (entities, attributes, relationships); at the logical level (relations, sets, records); and at the physical level (files, databases, indexes, pointers). It can also document procedures at the business level and at the application-program and clerical-procedure level. The data dictionary is therefore an inventory of the organisation's data resource.

A data dictionary system consists of a database (or set of files) containing the metadata, and a special database management system with some facilities that support data management such as data capture, indexing and reporting, and other facilities that support applications programs, such as the automatic generation of file definitions, validation routines and report layouts. The overall architecture of a data dictionary system is depicted in figure 6.

Most of the well-known database management systems now have an associated data dictionary system. For some products, such as IDMS, the two systems are closely integrated and the database management system makes use of data definitions and validation rules held in the data dictionary system. Other suppliers have designed their data dictionary systems to interface with a variety of database management systems. When selecting a data dictionary system care needs to be taken to distinguish between genuine systems, capable of documenting the full range of data and procedures described above, and those systems which are really an adjunct to a database management system. A list of data dictionary systems available in the late summer of 1982 is shown in figure 7.

User experience of data dictionary systems

A data dictionary system can be used in a variety of ways:

—To rationalise the data resources and the procedures which use them.

—To provide end users with information about the availability and characteristics of data.

—To document the results of data analysis and highlight omissions and inconsistencies.

—To provide data designers with information about data structures and uses.

—To promote co-ordination and data sharing between applications.

—To reduce routine program coding.

—To enforce systems standards.

—To analyse the impact of changes in business requirements on existing applications or databases.

—To assist with the identification of errors in applications.

Very few organisations use a data dictionary system so extensively. In our survey, about two-thirds of the organisations who responded claimed to be using a data dictionary system. Of these, 37 per cent had introduced their system during the previous 12 months, and 52 per cent had less than two years' experience. The data dictionary system was used most frequently to document programs for order processing, sales accounting, sales analysis and financial accounting applications, and to document record types (sometimes entity types as well) for customers, accounts and sales-order data. This evidence suggests that the use of data dictionary systems is growing quite rapidly but that the predominant use is at the technical, rather than the business, level.

Figure 6 Overall architecture of a data dictionary system

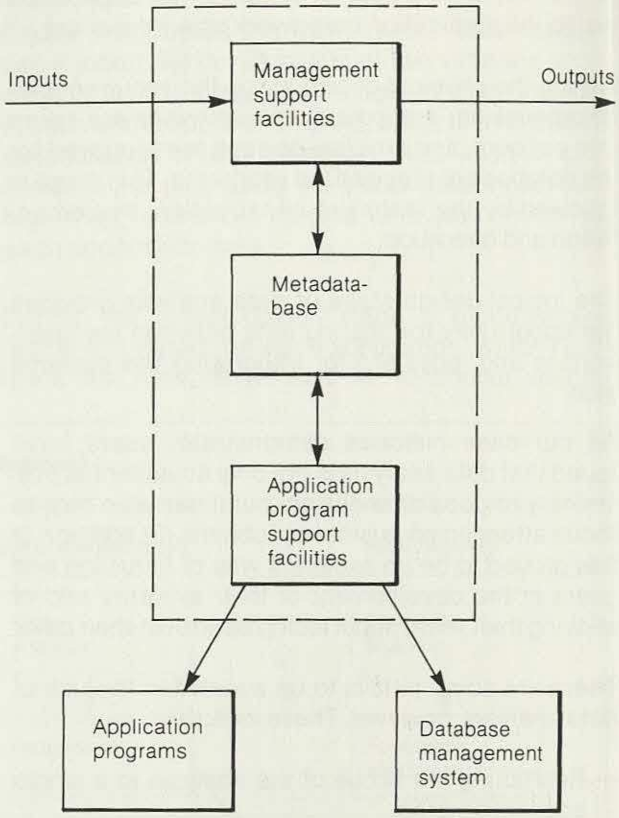


Figure 7 Data dictionary systems

Product	*Supplier*
Adabas DDS	Software AG
Byblos	Sligos
Data Manager	MSP
Data Catalogue-2	Synergetics
Datacom/DD	ADR
DCS	Haverley
DD/DS	Honeywell
DDS	ICL
DMS/DDS	Univac
Garde	SPI
IDD	Cullinane
IMS DDS	IBM
Pride-Logic	M Bryce & Associates
UCC-10	UCC

The two most frequently mentioned reasons for choosing a particular data dictionary system were:

—The opinion of the organisation's own technical staff.

—Compatibility with existing equipment or software.

We believe that the relatively limited use of data dictionary systems, in spite of their growing popularity, is due partly to the mistaken view that such a system is of value only to the database administrator, and partly to their lack of user-friendly facilities. It is often difficult to enter or amend data, there are only limited enquiry facilities and the output is generally in listed, rather than graphical, format. Interactive graphics would be of great value in generating and amending data-model diagrams, program charts and data-flow diagrams.

We expect that the facilities provided by data dictionary systems will improve slowly as a result of user pressure. We also expect that the emphasis will change from them being used as a passive documentation aid to being used as an interactive analysis, design and management aid.

DATABASE MANAGEMENT SYSTEMS

In chapter 1 we defined a database management system as a set of software tools designed to provide access to a computerised database; to enforce privacy, security and integrity of the data; and to provide special languages for describing the data relationships (a data description language) and for manipulating the data (a data manipulation language).

In our survey we found that 86 per cent of the organisations that responded were using a database management system and, of these, 60 per cent had more than two years' experience. It is reasonable to conclude that most large organisations are using a database management system, although their level of use varies greatly.

The first database management system, IDS, was introduced by General Electric in 1963 and most of the well-known products (IMS, IDMS, Adabas, Total etc.) were available by 1971. All such products have significant drawbacks, which include one or more of the following:

—Their use of one construct (such as the Codasyl set or the IMS parent-child pointer) to handle several different structuring concepts (such as program access paths, entity relationships, and entity inter-dependencies).

—Their procedural data manipulation languages, which process only one record at a time, making loops and iterations unavoidable, and limiting the scope for non-procedural query languages.

—Their lack of attention to operating efficiency or integrity constraints.

—Their inability to support different user views; that is, their lack of logical data independence.

—Their lack of physical data independence.

—Their poor end-user facilities.

—Their need for continual support from system programmers or database administrators.

It is important to recognise that these drawbacks are inherent in the basic architectures of the database management systems. In this sense they are similar to operating systems, and their suppliers all face the same dilemma: should they invest in a new product designed to overcome these weaknesses, and risk losing their captive markets, or should they build peripheral improvements into the products and counterbalance the basic weaknesses with strong marketing and support? Invariably the second option is chosen, and not without justification. Our survey showed, for example, that among the ten organisations in our sample that were using IMS the collective satisfaction rating for the product was only 62 per cent (compared with 80 per cent for Adabas and 73 per cent for Total). However, only two rated their experience with the product as worse than their expectation, and only one organisation would not use IMS to do the same job again.

In our survey we found that the three most frequently mentioned reasons for choosing a particular database management system were:

—Compatibility with existing equipment or software.

—Expected ease of use.

—The opinion of the organisation's own technical staff.

In Foundation Report No. 12 — Trends in Database Management Systems — we suggested that, once installed, the database management system would probably outlive the equipment and much of the other software in the installation. It would also influence considerably the type and quality of service offered to the rest of the organisation. Selecting the right database management system was therefore a critical managerial, as well as a technical, decision. We would now add that the choice is basically between one of the well-established, well-supported products with inherent weaknesses, and a newer product in which those weaknesses have been designed out. We discuss below some of the new and potential candidates.

Relational database management systems

A relational database is a store of data whose structure is expressed as relations — that is two-dimensional tables in which the rows represent all of the data describing one entity (called a tuple and analogous to a record in a conventional file), and the columns represent all of the values of one particular attribute (called a domain and analogous to a field in a conventional file).

A relational database management system has the ability to separate or recombine tuples and domains to form new relations, which gives conceptual simplicity and great flexibility in the use of the data.

There is currently a great deal of interest in relational database management systems. This interest arises partly from the simplicity and attractiveness of the concept; partly from their ease of use, especially by end users; partly from the ease with which suppliers can develop additional associated products; and partly for purely marketing reasons — no self-respecting product could now fail to describe itself as relational.

It is very likely that a standard for relational database management systems will be produced by a committee of the American National Standards Institution. In the meantime, IBM is rapidly establishing a de facto standard for a relational data manipulation language with its recently released SQL.

At the time of writing this report (late summer of 1982) few fully relational database management systems were available, but the number was increasing as their use was becoming more widespread. Figure 8 lists the relational database management systems available at that time.

The performance of a relational database management system needs be no worse than that of a traditional system for equivalent tasks. Performance problems may well arise, however, from the ease with which non-procedural languages may be used to specify tasks that would otherwise be virtually impossible.

Figure 8 Relational database management systems

Product	*Host computer*
Oracle	PDP-11, VAX
RQL/32	Perkin Ellmer
Ingress	
XQL	
System 38 DBMS	System 38
Omnibus	IDM with VAX
Noah	IDM with Datapoint
Encompass	Tandem
RDBMS	ICL

In addition to relational database management systems, relational data manipulation languages will become available for use with traditional systems, but their performance may suffer from the constraints of the underlying database management system.

The availability of relational software will increase the need for good data analysis and data design, because the power of the software will be dissipated unless the data are completely consistent in both value and meaning.

Database management systems for microcomputers

The growth in the use of microcomputers has led to the development of microcomputer-based database management systems. Many of these products have been put together with the minimum of resources or skills but a few are notable in that, apart from storage volume, they offer better facilities than those of the traditional database management systems.

For example, MDBS III is a Codasyl-based system produced in the United States by Micro Data Systems of Indiana. It costs about $3,000, runs in any C-compiler or CP/M, MP/M, Unix or RSTS environment, and uses Codasyl sets with variable-length records. Also, it supports many-to-many and recursive relationships and has a Codasyl data manipulation language. It also has a query language suitable for text handling and a report writer with statistical functions.

Figure 9 lists the microcomputer-based database management systems that were available at the time this report was written.

User-friendly relational systems for microcomputers and the so-called 'pocket' database will be available within a few years. Data management will be essential if these facilities are to be controlled through the use of common data. The alternative will be a proliferation of independent, inconsistent and ineffective local databases.

Database machines

A database machine is a processor dedicated to database management functions. This objective may be achieved in two ways — dedicated function and extracted function. In a dedicated-function database machine the software has been written specifically to handle database management, and

Figure 9 Microcomputer-based database management systems

MDBS III	Codasyl-based systems
Microseed	Codasyl-based systems
CCA Data management system	
Database Two	
DBMS	
DNA	
Global	
IDM	
KSAM 80	
MDBS	
Microfile	
Micro Rapport	
Midas	
Pascal KSAM	
Vulcan	

there is no distinction between the conventional operating system and the database management system. Examples of dedicated-function machines are IDM from Britton-Lee Inc., IDBP from Intel Corporation, and DBM from Digital American Computers Inc.

The alternative approach is the extracted-function machine. This is a conventional applications processor, with an operating system and a database management system that are dedicated to database management. An example is the Adabas database machine from Software AG in West Germany.

The advantage of the dedicated-function machine is that potential conflicts between an operating system and the database management system are avoided. As a result, the software is simpler and more efficient, and the non-essential functions of a conventional operating system are omitted. Furthermore, the database machine may be linked to more than one host computer, and these may be different types of equipment using different operating systems. This feature enables data to be shared between applications written for different computers.

The disadvantage of the dedicated-function database machine is that a software module is required to handle requests from the host computer, and to receive or deliver data. We believe that this disadvantage is outweighed by the advantages described above.

The advantage of the extracted-function machine is that it is much cheaper to develop than the dedicated-function machine. But it lacks the benefit of special software and it incurs the penalty of the host communication overhead. Consequently, we do not see much future for the extracted-function machine.

We expect to see database machines developing in three directions. First, the relational database machine would be a natural development, especially if it included parallel processing, making it particularly suitable for use with non-procedural languages. Such a machine would have the ease of use inherent in a relational database management system, coupled with fast processing from the specially designed equipment.

The IDM from Britton-Lee Inc. is a relational machine. It can handle up to 2,000 transactions per minute and a database of up to 32 gigabytes. Its cost, excluding disc storage, is in the range $50,000 to $150,000, which compares well with conventional database software.

The second development that might occur is a further migration of the functions conventionally resident in the mainframe host computer. Data storage, transaction update and access, data validation, security and integrity could all be handled by a database machine. Screen management, user-friendly query management and transaction management could all be handled locally by intelligent terminals. The only functions remaining in the host would be mass data entry, batch processing and processor-dependent work such as scientific computing. The third of these functions is already migrating onto special-purpose machines, and it might make sense to extend the abilities of the terminals and database machine to handle the other two tasks. The host would then remain as a link with the past, running existing application software until it was redeveloped to take advantage of the new environment.

A third development will be the introduction of database servers, connected to local area networks. The servers will be able to service the data needs of the terminals within the network, not only for locally stored data but also, if necessary, for data stored on other networks or at a central processing site.

Distributed database management systems

A distributed database management system supports a single unified view (a conceptual schema) of data that are distributed between processors at different geographical locations. Application programs do not need to know where the data are stored.

The problems to be solved in designing a distributed database management system are:

—Optimising the location of data storage to minimise the transfer of data across the communication network.

—Optimising the sequence in whch data are retrieved in response to an application request, to minimise the transfer of data.

—Avoiding or resolving deadlocks arising from updating requests from different locations.

—Co-ordinating the recovery, after failure, of transactions which may be in many different states in different parts of the network.

—Achieving consistency of update, especially where data are replicated.

—Interfacing different processors or terminals in different parts of the network.

A number of suppliers are developing distributed database management systems. These suppliers include Cincom Systems Inc., Cullinane Database Systems Inc., Infodata Systems Inc. and Software AG. To date they have avoided many of the above problems by offering so-called distributed database management systems which have only limited facilities. We believe that none of the systems permits the applications programmer to be unaware of the location of the data. Full geographical independence is not yet available.

We recommended in Foundation Report No. 18 — Distributed Processing: Management Issues — that the best solution to the need to make data available at several locations was to transfer data in bulk for either local enquiry or central update. We still believe that this recommendation is valid, but the co-ordination of such transfers is a data management function that should not be biased towards any one application.

NON-PROCEDURAL LANGUAGES

Non-procedural languages are a specially important set of tools. They differ from conventional high-level programming languages in that they state what has to be done rather than how it is to be achieved. Their structure is designed for handling many records of the same type concurrently.

We addressed their use and value, especially to end users, more fully in Report No. 30 — End-User Computing. In the context of data management their importance lies in the ease with which they can be used to access a database or data dictionary system, and so to increase the value of these tools. This will be especially true when relational database management systems and their associated query langages become available. In our opinion, the full benefits both of database management systems and of data dictionary systems cannot be achieved unless a non-procedural language is used to satisfy ad hoc demands for information.

DATA DESIGN AIDS

There is a small but growing family of software tools that have been developed to automate some of the activities of data analysis and data design. Two such products (Data Designer and Data Planner) are available from DMW Group Inc. A similar product (Design Manager) is available from MSP Inc. and complements MSP's data dictionary product, Data Manager.

Data Designer and Design Manager are used to synthesise one logical data model in third normal form from a combination of any number of local views, which may be input forms, clerical records, computer listings, file structures, etc. Any redundancies or inconsistencies are highlighted. The output is a set of logical attributes and relationships, with the overall usage frequency and required response time summarised for each attribute.

These products currently support only a small part of the analysis and design process. We regard them as precursors of much more powerful tools that will become available within the next few years.

FUTURE DEVELOPMENTS

We referred earlier to the dilemma faced by the suppliers of software tools who have a well-established product with a large captive user base, but which may contain basic design weaknesses. We suggested that they will attempt to counterbalance the weaknesses by providing compensating benefits and strong marketing and support. The kinds of action that suppliers are taking include:

—Developing closely integrated sets of data management tools, such as ICL's IDMS and DDS products.

—Developing interfaces between established products and quasi-relational products; for example, IBM's SQL/DS linking DL/1 databases with the SQL relational language.

—Integrating teleprocessing monitors with database management systems; for example, IBM's IMS DB/DC.

—Developing distributed versions of established database management systems; for example, Adabas.

—Developing application packages based on existing database management systems; for example, Cincom, the supplier of Total.

—Developing text-handling features for existing database management systems.

All of this activity will benefit users if it leads to efficient interfaces and to the removal of redundant logic, but not if it is merely repackaging existing products for marketing reasons.

SUMMARY

The conventional emphasis in the application of database technology has been to focus first on the database management system, then on the query languages, with the data dictionary systems and data analysis coming a poor third. We suggest that the correct emphasis should be in the reverse order. First, data analysis should be recognised as a valuable aspect of computer system development and an essential aspect of data management, and the data dictionary should be viewed both as a vital tool of data analysis and for holding an inventory of the organisation's data resource. Next query languages, especially non-procedural languages, should be recognised as a crucial means of accessing the data for the programmer and the end user. Finally, a database and a database management system should be used only if the circumstances justify them.

CHAPTER 5

THE ORGANISATION OF DATA MANAGEMENT

Chapter 1 identified the importance of data as a corporate resource and argued that they deserve to be managed formally and deliberately like other corporate resources. This chapter discusses the ways in which those activities and responsibilities may be organised, and highlights some of the difficulties that must be overcome in setting up effective data management.

POLICY ISSUES

Many organisations still find it difficult to accept that capital or staff are corporate resources (rather than local resources) and that they should be exploited by sharing between many users.The concept of managing data as a corporate resource will therefore not be accepted easily. Before it is accepted, two prerequisites will have to be satisfied:

—Top management in general and end users in particular should understand why data are an important resource.

—Top management should accept a commitment to promote the introduction of data management.

If these prerequisites are satisfied, then policies are required that indicate the organisation's intentions regarding:

—The allocation of responsibility and accountability for data management.

—The formulation of a strategy for the development of the data resource.

—The ownership and sharing of data.

—The scope of data management.

Allocating responsibilities

The allocation of responsibility and accountability for data management is a classic dilemma. Some organisations are unwilling to commit themselves totally to data management but claim to be making steady progress by applying the tools and techniques to individual applications. They believe that the benefits of data management will be demonstrated by the initial applications, and that it is not necessary to set up a formal data management organisation from the beginning. They are confident that their top management can be persuaded to approve such an organisation once they see the initial benefits.

Other organisations argue that the success of data management depends on solving the political and organisational difficulties created by data sharing, rather than on applying particular tools or techniques. They believe that the full benefits of data management will never be achieved until responsibilities and accountabilities are assigned formally, and that the organisation structure must therefore be set up as part of the initial investment.

It is certain that no organisation will achieve the full benefits of data management without allocating these responsibilities, but the choice of whether to do this sooner or later is a matter of policy.

Formulating a strategy

A strategic plan for developing the data resource is required because, inevitably, this development will be a long-term process involving many different activities. Without a strategic plan the ultimate objective may become obscured by short-term requirements, and the justification for the investment in data management may be undermined. The lack of a strategic plan also means that there will be no yardstick against which to measure progress or to detect the need for corrective action. Different organisations may well formulate different strategies. What is certain is that no organisation will achieve the benefits of data management by default — as a matter of policy a strategy is essential.

Sharing data

Ownership of data is a difficult issue. In theory data are owned by the organisation, with responsibility for defining, creating and maintaining them delegated to individual departments or managers. In practice data are regarded as personal property and, inevitably, this leads to conflicts over their use and inhibits data sharing. One of the functions of data management is to break down these inhibitions. This will be possible only if top management make it clear that the policy of the organisation is for data to be shared, and that it is the prime responsibility of those in data management to achieve this.

Defining the scope of data management

The scope of data management is a less contentious issue that can generally be resolved on a practical basis. It has two aspects. First, the organisational scope for data management must be defined. A balance must be struck between defining this scope too widely (for example, to include the whole organisation) in order to maximise the potential for data sharing, and too narrowly (for example, a single application area such as production management or management accounting) in order to simplify the data management strategy and minimise political or organisational problems. A strong argument for making the scope as wide as possible is that the conceptual data model should not be constrained by the current organisation structure. The need to amend data models as a result of organisation changes should be avoided.

The second aspect is to define the scope in terms of the type of data that are subject to data management. At least three boundaries can be identified, the narrowest being those data which are held in computer databases, the next being all data stored on a computer, and the widest being all corporate data whether stored on a computer or not.

DEFINING THE RESPONSIBILITIES

Any organisation intending to introduce data management must accept that, sooner or later, the responsibilities must be defined and focused on one individual, who then becomes accountable for the control and cost-effective use of the data. Our survey showed that, in practice, there has been considerable confusion in defining those responsibilities and in locating the individual within the organisation structure. Conceptually, the responsibilities divide into two groups.

Data administration responsibilities

The first group of responsibilities is concerned with the organisation in general and with its use of data. These responsibilities include:

—Creating an understanding of data and of its importance.

—Reinforcing the commitment of top management to data management.

—Formulating the data management strategy.

—Resolving disputes about the ownership of data.

—Defining standards for the description of data.

—Selecting and introducing tools and techniques.

—Creating and maintaining the corporate data model.

—Defining and enforcing controls over access to data, security, recovery and integrity.

—Recognising and promoting opportunities for data sharing.

—Designing and maintaining the applications framework.

—Monitoring the cost-effectiveness of data handling systems and advising where redundant data could be avoided.

—Recognising and controlling trade-offs between long-term and short-term considerations, such as the desire to retain flexibility and minimise system maintenance costs versus the need to maximise performance or minimise the cost of system development.

—Advising on the impact that changes in business requirements or data definitions or data handling procedures could have on other parts of the organisation's systems.

In summary, these responsibilities are aimed at achieving data sharing, logical independence and the cost-effective use of data. Note that none of the responsibilities is computer-dependent, although many would have enhanced significance in a computing environment. This group of responsibilities is often referred to as data administration.

Database administration responsibilities

The second group of responsibilities within data management concerns the storage and handling of data. Although they could apply in a non-computerised environment these responsibilities would be relatively trivial in that context. They are most easily defined in a database environment where they are commonly referred to as database administration. These responsibilities include:

—The logical and physical design of individual databases, derived from the corporate data model.

—Advising systems analysts and programmers on the use of databases.

—Avoiding application requirements that are valid logically but impossible practically.

—Reconciling conflicting requirements, especially where new applications might create problems with existing applications.

—Implementing controls and procedures to satisfy the specifications for access, security, recovery and integrity laid down by the data administrator.

—Monitoring the usage and performance of the database and amending its physical structure to improve performance where necessary.

—Providing technical support for users of software tools such as database management systems, data dictionary systems and data design systems.

In summary, the responsibilities of database administration are aimed at achieving physical independence, and the cost-effective storage and handling of data.

ORGANISING THE RESPONSIBILITIES

What then should be the relationship between data administration and database administration, and where should they be located within the organisation? The principal considerations, some of which are in conflict with each other, are:

— Data administration ideally should be a high-level independent department, like the personnel department or the treasury. Such status is essential to emphasise the importance of data and to give the data administrator the required authority.

— Data administration should be outside the computing department since its orientation is towards the organisation and its use of data, and most of the data resource is not stored on the computer.

— Database administration should be separate from application development, with adequate authority to ensure that the applications framework and database standards are followed.

— Database administration should be a part of data administration because its role is to implement policies and standards laid down by the data administrator.

— Database administration should be part of the computing department because all of its activity involves computerised tools and techniques.

Clearly, it is not possible to satisfy all of these requirements and, in practice, data administration (if it exists at all) is usually located within the computing department. This is understandable because it is generally the head of computing who recognises the need for — and the political significance of — data management. This arrangement also helps to disperse any friction between traditional computing staff and the newcomers in data management. However, it also reinforces the erroneous impression that data management is really an aspect of computerised databases, and so data administrators have to avoid being reduced to the role of database technicians.

Data administration

In the majority of organisations that responded to our survey, data administration was located within the data processing department. Most of them had appointed a person to be responsible for data management, of whom 21 per cent reported to the data processing manager, 39 per cent reported one level below that and 27 per cent reported at two levels below the data processing manager. When we compared organisations which we regarded as having established data management with those which had not, a clear shift of responsibilities to the data administrator and database administrator was evident. Planning responsibilities had shifted at the expense of systems managers and long range planners; system development responsibilities had shifted at the expense of business analysts, application project teams and systems programming staff; and operational responsibilities had shifted at the expense of computer operations and systems support staff.

The style with which the data administrator discharges his role depends upon his degree of authority. Three types of role may be distinguished, corresponding to three levels of authority.

Missionary role
The data administrator is responsible for encouraging the sharing of data, the use of data management tools and techniques and the adoption of standard names, formats and coding systems, but he has no authority to impose his views. This is often the initial role.

Enforcement role
Policies and a strategy for data management are determined by someone other than the data administrator, but he is responsible for enforcement and has the authority to resolve disputes.

Manager role
The data administrator is fully responsible for all aspects of data management and is accountable for its success.

Database administration

Database administration is always located within the computing department. In some cases the responsibilities of database administration are enlarged to include all computer-stored data. The rationale is to reduce duplication of effort and to encourage the sharing of data by removing from the application development teams all direct responsibility for data files. This principle can be applied even when there is no database.

The database administrators in practice find themselves responsible to a variety of managers within a computing department — technical support, systems development or head of data processing. Ideally, the database administrator should report to the data administrator, if the latter is located within the computing department.

SKILLS REQUIRED

Our survey suggested that data management staff nearly always have a data processing background. It is desirable that database administrators should be skilled application designers or systems programmers. It is less obvious that data analysts should have previous experience as systems analysts, even though the aptitudes for observation, analysis, synthesis and thoroughness are common to both tasks. The skills of tact, diplomacy and political awareness are of greater importance for data administration than for computer systems development. It is therefore wise to look beyond the computing department when selecting data administration staff.

DATA MANAGEMENT STANDARDS

One of the principal responsibilities of data management is to define standards that will enable data to be shared and used most effectively. Typically, standards are required in the following areas:

- —Data descriptions: for defining data classes, formats, attributes, relationships.
- —Techniques: for describing analysis and modelling techniques and their associated diagram conventions and documentation.
- —Tools: for using data dictionary systems, database management systems and data design aids.
- —Programming languages: for processing data.
- —Security and integrity: for defining validation rules, and access or recovery procedures.

CENTRALISATION AND DISTRIBUTION

By its very nature, data management is a centralising influence because it is concerned with standardising and sharing. However, many organisations have distributed much of their formal data processing and, as we showed in Report No. 30, end-user computing is growing rapidly.

Data administrators should not ignore this trend. Data processing staff at distributed computing centres, and end users, are carrying out activities similar to those of staff located at the main computer centre, and they deserve similar support. Also, data management provides one effective way of controlling unauthorised system development activity because it can control access to, and use of, corporate data.

In some large organisations local data administrators have been appointed who are responsible to, or co-operate with, the corporate data administrator. Their role is to assist the local data processing staff and groups of end users to make the best use of local or central databases, and to exercise control over their use of data.

SUMMARY

In this chapter we have discussed the ways in which data management activities may be organised, and we have highlighted some of the difficulties to be overcome before an effective data management function can be established. But there is another difficulty that has to be overcome before the organisation of data management can even be considered — the difficulty of justifying data management in the first place. What does data management cost? What are its benefits? How does an organisation cost-justify data management? We address these questions in the next chapter.

CHAPTER 6

JUSTIFYING DATA MANAGEMENT

In this chapter, we first discuss the situations in which data management may or may not be relevant, then summarise the main costs and benefits associated with the various elements of data management, and finally suggest how the case for data management can be put convincingly to top management.

THE RELEVANCE OF DATA MANAGEMENT

We believe that the majority of organisations can benefit from adopting data management. Its relevance appears to depend neither on the size and activity of the organisation nor on the types of application for which computers are used.

Data management is most likely to bring benefits if the following conditions apply:

—The organisation is of average or above-average complexity, with many data relationships and business functions. (Companies undertaking data analysis often identify a much larger number of relationships than they had expected.)

—The organisation is large, even if not complex, and so has many data flows and consequent communication problems.

—Data are or could be shared between several users or activities.

—A management information system is needed which must be flexible and must respond to changing business needs.

—The major computer systems are distributed or decentralised. Such systems often lead to steadily increasing inconsistencies and communication difficulties.

—There is no framework within which systems are to be developed.

—There is more than one database management system in use.

—There is a proliferation of systems developed by local users outside any systems framework.

On the other hand, data management may not be relevant if:

—The organisation is small or simple or very stable.

—The existing systems are simple and effective and are unlikely to need replacing.

—The existing systems and the underlying business processes are well understood and documented.

—The rate of change of systems and data is low and there is little need for system maintenance.

—There is little pressure for further system development.

COSTS AND BENEFITS OF DATA DICTIONARY SYSTEMS

In summary, the main benefits arising from the use of a data dictionary system are:

—That it will increase the availability and value of the data resource.

—That it will reduce the system life-cycle cost (even in the short term).

—That it will make possible the development of complex integrated systems.

—That it will support the enforcement of standard methods.

We believe that a data dictionary system would deliver benefits to any organisation that either is highly dependent on the availability of good information or has an appreciable data processing activity.

The additional costs incurred by using a data dictionary system are those of installing, loading and supporting the system.

In our survey the most frequently mentioned benefits of using a data dictionary system were:

—For the organisation in general: availability of metadata, quality of metadata and management control.

—For the data processing department: consistency of metadata and quality of documentation.

We asked the survey respondents to rate on a scale from 5 to −5 the benefits or difficulties they expected from using a data dictionary system, and also to rate the benefits they actually experienced. (Zero meant no impact; 5 meant extremely beneficial; −5 meant extremely difficult.) The aggregate of the responses is summarised below:

- —For the organisation, the expected benefits were rated at 3.2 and the benefits actually experienced were rated at 2.4.
- —For the data processing department, the expected benefits were rated at 2.7, compared with 2.1 for the benefits actually achieved.

These benefit ratings are surprisingly low, but they need to be compared with the satisfaction rating of 3.2 (on a scale of 0 to 5) and a product loyalty of 88 per cent.

COSTS AND BENEFITS OF DATA ANALYSIS

In summary, the main benefits arising from the use of data analysis are:

- —That it provides the basis for a disciplined approach to systems development at both strategic and application levels.
- —That it is an essential technique for analysing and documenting the data resource and for defining standards for naming and defining data.
- —That it is an essential precursor to data sharing and effective data design.
- —That it can easily be applied by end users and helps to involve them in the development of their systems.

The additional costs of adopting data analysis are not only the resource costs of performing a task that would previously have been omitted, but also the political difficulty of persuading top management that earlier methods were defective and yet another new technique should be introduced.

In our survey the satisfaction rating amongst those organisations that had used data analysis was 3.7 (on a scale of 0 to 5).

COSTS AND BENEFITS OF DATA ADMINISTRATION

In summary, the benefits of appointing a data administrator to manage the data resource are:

- —That he will be the essential agent of change, smoothing the introduction of data management and helping to resolve the inevitable political difficulties.
- —That he will formulate and maintain an overall strategy for data management.
- —That he will ensure that data sharing, logical independence and cost-effective use of data are achieved.

The additional costs associated with data administration are not only the resource costs of the data administrator and his staff, but also those that may be caused by the political difficulty of making the act of faith to appoint him in the first place. Amongst those organisations which responded to our survey and which we regarded as having introduced data management, the average annual cost of data administration was about 7 per cent of the data processing budget.

COSTS AND BENEFITS OF DATABASE AND DATABASE MANAGEMENT SYSTEMS

The arguments for storing data in a database and using a database management system have been well rehearsed for many years and were set out in detail in Foundation Report No. 12. In summary, the main benefits of the database approach are:

- —That it will make the corporate data resource independent of both current uses and current storage methods, and so will make that data more useful, especially for management-control and strategic-planning purposes.
- —That, in the long term, it will reduce the life-cycle cost of systems.
- —That it will reduce the time required to develop new applications (especially ad hoc data retrievals) and so deliver their benefits earlier.
- —That it will increase the security and integrity of the data resource.

We would not claim that a database is a better solution than conventional files in every situation. But if data may need to be shared and if security and integrity are important, then a database should be considered unless there are overwhelming performance penalties.

The additional costs associated with the database approach are those of installing and supporting the database management system. There are also the costs of data analysis, which we discussed above.

In our survey, the most frequently mentioned benefits of using a database management system were:

- —For the organisation in general: availability of data and quality of data.
- —For the data processing department: operational reliability and data consistency.

As with data dictionary systems, we asked the survey respondents to rate the expected and achieved benefits and difficulties for database management systems on a scale of 5 to −5. The aggregate of their responses was:

—For the organisation: expected benefit rating of 3.0, compared with 2.4 for benefits actually achieved.

—For the data processing department: expected benefits of 2.0 compared with 1.4 for benefits actually achieved.

Again, these ratings reveal a surprisingly low expectation and a disappointingly low experience of benefit arising from the use of a database management system. These ratings should be compared, however, with a satisfaction rating of 3.4 (on a scale of 0 to 5), which is almost the same as in our 1979 survey, and a product loyalty of 89 per cent.

PERSUADING TOP MANAGEMENT

The overall picture to emerge from our discussion of the costs and benefits of the individual elements of data management is of real, quantifiable costs being incurred before any benefit is gained. Moreover, the benefits are mainly unquantifiable, sometimes intangible and are achieved well after the initial investment. It is clear, therefore, that the case for data management will not be self-evident. Nevertheless it may be possible to find cases where the costs and benefits can be quantified. Examples are:

—The cost of recreating essential files should they be destroyed.

—The cost of faulty decisions due to poor data.

—The cost of continual investment in poorly designed information systems.

—The cost of reconciling incompatible systems that could and should be integrated.

—The cost of maintaining duplicated data that could and should be shared.

Any of these examples may provide supporting evidence in justifying data management but they will rarely prove the case. Inevitably, those who would wish to introduce data management must persuade their top management that this is the right thing to do. Three lines of argument can be used to convert top management to this view.

Logic

We believe that any person with a reasonable understanding of the role of data within information systems would not dispute the benefits for the various aspects of data management that we have described in this report, even though they might dispute the balance of advantage.

But to be persuaded of the need for data management by logical arguments alone does to some extent require an act of faith. There are many examples in the history of data processing where concepts advocated by early pioneers have subsequently become accepted practice. Originally they had no tangible justification, but were seen to be the only logical way of doing the job. Structured programming is a good example.

Provided top management has a reasonable understanding of information systems, we suggest that the logic for introducing data management is compelling, and that there is no need to wait until data management is commonplace before being convinced of its benefits.

Experience

The logical argument will be greatly reinforced if reference can be made to other organisations who already have adopted data management with clear advantage.

One of the purposes of this report has been to describe the experience of organisations with data management. As our case histories showed, there are organisations who have adopted data management, and can point to the benefits and are convinced they took the correct decision. But in our research we also found a disappointingly high number of organisations who are adopting data management in a rather piecemeal fashion or who are no longer convinced that they took the right decision.

Regrettably, this line of argument is not yet totally convincing.

Good practice

Data management (and a database management system in particular) undoubtedly provides benefits for the data processing department. But these benefits alone will be insufficient to convince top management of the need for data management. The final argument is that, in a complex and competitive environment, good information is essential and that the data from which that information is derived is a corporate resource. It is therefore one of the duties of a responsible management to manage that resource.

Taken together, we believe that these three lines of argument will be sufficient to persuade top management to take the essential first step and make a commitment to introducing data management. The next chapter highlights the actions that need to be taken once that commitment has been made.

CHAPTER 7

GUIDELINES FOR ACTION

In this final chapter we provide some guidelines for organisations intending to introduce data management.

PLANNING FOR DATA MANAGEMENT

Data management will never become established by default. It is essential that a long-term strategic plan be prepared in order to introduce data management and achieve its benefits. In our research we found several organisations that had set out with the best of intentions but had then lost their way through lack of a proper plan, or lack of senior-management commitment.

Data management should be introduced in two stages, as follows:

Stage 1: Preliminary work

—Ensure that top management has the necessary appreciation of data as a resource and is prepared to make a commitment in principle to data management.

—Select a suitable methodology for system development and also a data dictionary system. The methodology should give adequate emphasis to data, and it should be selected together with the data dictionary system because each will support the other.

—Define and document the standards and working methods for all aspects of data management.

—Apply the data analysis parts of the methodology to a major new system development project. This will require a limited education programme for end users and technical staff.

—Select a database management system (if required).

—Assess the impact of data analysis and the data dictionary system on the project and identify the benefits achieved.

—Obtain a substantive commitment from top management for the full introduction of data management.

Stage 2: Establishing data administration

—Appoint a data administrator, ensuring that his unit is staffed adequately and has clear terms of reference, and that other peoples' terms of reference are adjusted accordingly.

—Set up an education programme for end users and data processing staff.

—Apply data analysis at the corporate level and develop a high-level corporate data model.

—Prepare a strategy for the incremental development and subsequent integration of subject databases, based on the corporate data model. Prepare an applications framework.

—Select any additional software tools.

—Apply the full system development methodology to all development projects now within the applications framework.

IMPLEMENTING THE PLAN

The approach outlined above leads to a gradual increase in the penetration of data management and of the commitment to it. We recommend that the tools, techniques and standards be well established before too much analysis and development work is carried out. We do not recommend that a detailed data analysis be carried out over the entire organisation. An initial high-level data model is sufficient, which can be progressively elaborated to greater levels of detail.

Evolving databases

If the data management strategy does lead to the development of databases it is important that they should evolve gradually. Initially, physical databases should be separate (but they should be designed so that they can subsequently be merged), and their use should evolve from batch processing to interactive update and retrieval. This gradual evolution is perfectly feasible provided that all of the physical databases are based on a common schema and that a standard data description language and data dictionary system are used throughout the development.

Anticipating changes

It is also important to recognise the changes brought about by data management. The changes in the techniques and the software tools will be very obvious but they should not obscure the fact that project management methods may also need to change. The balance of tasks will be different, and so costing and timing estimates will need to be adjusted.

Choosing the first project

The choice of the first project is particularly important because data management as a whole will be judged on the results. Ideally, the application should be clear-cut with definable boundaries and a minimum of interaction with related applications. It should neither be so large as to be unmanageable, nor so small that an application-oriented view is inevitable. It may well be in an area that is poorly understood and is currently causing problems, so that the benefits of data management will be readily apparent. The application area must be the responsibility of managers who appreciate the potential benefits of data management and who will actively support the analysis work. Experience shows that the area chosen is usually one that is likely to be supported by a database management system, because the new analysis techniques can then be applied to database design.

Planning the transition period

The implementation plan for data management must make full allowance for the transition from the existing environment to the new one. For most installations it will be many years before all existing systems can be changed to accommodate the new approach. During this long transition period there will be data that are not shared, or that may not come within the scope of data management, and there will be areas where the new methodologies are not being applied. This transition period will not be made easier by divorcing data management from system development. Quite the reverse is true, because data management must be seen as being a service to system development. Each new project and each new data area must be evaluated against the implementation plan, and the plan should make it clear whether or not the data management approach should be followed.

Using the data dictionary

All projects following the data management approach should be documented in the data dictionary. This does not imply that all existing systems need to be documented also, only that duplication of data between new and old systems needs to be carefully controlled.

HUMAN ISSUES

The introduction of data management requires changes in attitudes and working habits, and is likely to lead to resistance from established staff and possibly friction between them and those who represent the new approach.

The friction arises first at a technical level, because many of the newly assigned responsibilities of database administration would otherwise have been carried out by application development teams. Nevertheless, if data are to be shared, it is essential that these responsibilities be separated from those of application development. One of the advantages of locating database administration within the data processing department is that a close working association between the two groups of staff will help to dispel suspicion and friction.

Resistance to data management can also be expected from users, who see their ownership and control of data threatened. We have already mentioned that involving users in data analysis has proved a very effective way to remove their fears. We have also found it useful to establish a working group of users and data processing staff to analyse the existing problems of data and systems development, to explore the potential benefits of data management and to set objectives for improvement.

The duties of this group may well include the supervision of a pilot project and an assessment of its success. Data management must be seen, by all the many people who will be affected, to be the right solution, and not to be a threat to their positions. They must also see that it has the backing of top management. The greatest difficulties arise when data management appears to be imposed, when there are no group discussions or presentations and when the objectives and implications remain unclear.

Resistance from senior managers stems from a general lack of understanding. They may see the topic either as a detail of no concern to them, or as common sense which is already being applied, or as empire building by the data processing function. They are also typically reluctant to agree to an organisational change that may be disruptive, and whose benefits may be intangible and not fully proven. Yet we stress again that data management cannot succeed without the support of senior management. It is often easier for external consultants to achieve the necessary understanding by senior management than it is for the company's own staff.

EDUCATION AND TRAINING

Education and training is another vital aspect of introducing data management. Obsolete ideas have to be replaced and new attitudes and techniques explained. An education programme is likely to include different sessions for the various groups of people involved.

Senior management undoubtedly need an appreciation of what data management is intended to achieve, the difficulties that need to be overcome, the policies that will have to be established and the contribution that will be expected of them.

A similar session will be needed for end-user management, concentrating on how they will benefit, on the realities of data sharing and on the participation that will be needed from key people during the data analysis stage. Additional sessions will be needed if end-user facilities are to be provided for access to shared data.

The data analysts, and selected users who will join the analysis teams, will need a thorough grounding in the methodology. This is best handled by workshop sessions based on the users' environment followed by on-the-job training.

The systems analysts or the application project teams will also need some training in the methodology. They must be convinced of the need to adapt to the new approach, and they must understand the details they have to supply to the data management team. They must also be able to apply the data analysis techniques themselves when this could be useful.

The database administrator will need training in data analysis techniques, and in how to use the results of the analysis in the design of data structures for the selected database management system.

The training for programmers will be product-oriented as they are the group least affected by a data management approach. They may have new languages to learn, and with most database management systems they must be aware of the performance implications of the way different language statements are used.

With the exception of senior management all these groups will probably be concerned with the data dictionary. The relevant dictionary facilities and standards should be regarded as an integral part of each training session.

IMPLEMENTING A DATA DICTIONARY SYSTEM

The data dictionary system can become a useful tool at the application-development level as soon as it is introduced. It will not be able to play its full part in data management, however, until a significant proportion of the data resource is recorded. The data management strategy must therefore address the problem of fully implementing the data dictionary system.

Three options are available. First, any data stored in existing databases can be documented (perhaps automatically, with any duplication or inconsistencies being removed manually), and additional data added as each new application or data class is developed. Second, data involved in one particular application area can be documented irrespective of whether they are already stored in a database, and then the scope of the data dictionary system can be extended to cover other applications. The third option is to begin by documenting entities and business procedures in the organisation as part of the data analysis exercise, and then gradually work through, area by area, to recording details of physical storage and programs.

Whichever option is selected, the data administrator must recognise that maintaining data within the data dictionary system is just as important as adding new data.

MISTAKES TO AVOID

During our research for this report we identified several mistakes which commonly undermine the introduction of data management.

Level of responsibility

Responsibility for data management is allocated at too low a level in the organisation structure. The data administrator therefore lacks authority and the subject is regarded as of little importance.

Unclear responsibilities

The data administrator is not given clear responsibilities, and the suggested responsibilities overlap with those of other functions such as computer operations, system development and organisation and methods.

Conflicts of interest

The data administrator is given clearly defined responsibilities, but corresponding changes are not made to other related functions. As a result, there are continual conflicts of interest and data management is often perceived as a rival to systems development instead of being a service.

Poor communications

Formal means of communication in both directions are not established with other related functions.

Implementation sequence

Data management is formally established only after data analysis, a data dictionary and a database management system are already in use.

Applications framework

Data management is not given responsibility for the applications framework, which either does not exist, or which begins to conflict with the models that data management must maintain.

Incorrect focus

The data administrator concentrates exclusively on analysing existing data and fails to take a wider business-oriented view.

But perhaps the biggest mistake of all is the failure to involve top management. In some organisations this is not so much a mistake as a difficulty that is not overcome. Data sharing does raise a host of political problems, and these are more easily dealt with if top management has indicated its determination that they will be resolved. With support from the top, the necessary user involvement can be assured, and a corporate view of priorities can be taken. There is also more chance of avoiding rival incompatible data structures and continued redundant development.

Data analysis almost invariably reveals anomalies, inefficiencies and organisational issues that management needs to ensure are addressed. If data have been recognised as a vital corporate resource then they deserve to be planned from the top.

CONCLUSION

This report set out to answer three questions:

—Have any organisations successfully adopted data management?

—What have been their experiences?

—How can an organisation justify data management?

During our research we encountered organisations that have adopted data management, and can point to the benefits and are convinced they took the correct decision. But we also found a disappointingly high number of organisations which are adopting data management in a rather piecemeal fashion or are no longer convinced that they took the right decision.

So, the answer to the first question is a qualifed yes. From those organisations who have successfully adopted data management, we have identified the practices that have contributed to their success and distilled them in the later chapters of the report. But we have also drawn out the lessons to be learnt from those organisations that have been less than successful. These lessons were highlighted in chapter 7 as the most common mistakes to avoid.

For most organisations, the most difficult question to answer will be how to set about justifying data management. As we indicated in chapter 6, the costs associated with data management are real and quantifiable, and they need to be incurred well before any benefits are gained. And the benefits themselves are mainly unquantifiable and sometimes intangible.

The report has shown that an organisation that wishes to introduce data management needs to make considerable investments in terms of manpower, training, education, and tools and techniques. We believe that the benefits to be gained from adopting such a policy are substantial. Indeed, we would say that data management should be an essential part of any organisation's overall systems strategy. But for data management to be successful it requires more than tools and techniques and expertise. It requires a commitment throughout the whole organisation to the concept of data management. In turn, this requires that the senior management of the organisation is convinced of the need for and the benefits of data management.

We believe that this report has identified the key issues of data management and has presented them in a way which should assist senior managers in understanding the role of data management in a modern organisation.

BIBLIOGRAPHY

1. Codd, E. F., *Relational database — a practical foundation for productivity,* IBM Research Laboratory, November, 1981.
2. Curtice, R. M. and Dieckmann, E. M., "A survey of data dictionaries", *Datamation,* Vol. 27, No. 3, March, 1981, p. 135-158.
3. Fillenson, M. L., *Data administration survey report,* IBM Systems Research Institute, 1981.
4. Flavin, M., *Information modelling,* Yourdon Inc., May, 1980.
5. Guide International Corporation, *Data administration methodology,* 1978.
6. Hubbard, G. U., *Computer assisted database design,* Van Nostrand Reinhold, 1981.
7. *Business systems planning — information systems planning guide,* IBM, July, 1981.
8. Kent, W., *Data and reality,* North-Holland, 1981.
9. Lomax, J., *Evaluating the database approach,* National Computing Centre, 1979.
10. Martin, J., *Information engineering,* Savant Institute, 1981.
11. Martin, J., *Strategic data planning,* Savant Institute.
12. Murgenberger, H., "Database administration — experience from a European survey", *Proceedings of the European Conference on Evaluation and Implementation of Database Systems,* September, 1980.
13. Olle, T. W., *Information system design methodologies — a comparative device,* North-Holland, 1982.
14. Palmer, I. R. and MacDonald, I. G., *System development in a shared data environment,* 1982.
15. Rock-Evans, R. and Palmer, I. R., *Data analysis,* IPC Business Press Ltd., 1981.
16. Ross, R. G., *Data dictionaries and data administration,* AMACOM, 1981.
17. Schneider, H. J. and Wasserman, A. I., *Automated tools for information systems design,* North-Holland.
18. Share Inc., *Data administration project — comparative analysis of tools and methodologies,* September, 1981.
19. Share Inc., *Data administration project — position paper,* December 1980.
20. Tozer, E. E., "Is distributed database yet mature enough to consider?", *Proceedings of the 22nd European Guide Conference,* May, 1981.
21. Weldon, J. L., *Database administration,* Plenum, 1981.

Butler Cox & Partners Limited
Morley House, 26-30 Holborn Viaduct, London EC1A 2BP
☎ 01-583 9381, Telex 8813717 LNCO

Belgium & The Netherlands
SA Butler Cox NV
Avenue Louise – 479 – Louizalaan,
Bte – 47 – Bus,
Bruxelles 1050 Brussel
☎ (02) 647 15 53, Telex 61963 BUTCOX

France
Butler Cox SARL
Tour Akzo, 164 Rue Ambroise Croizat,
93204 St Denis-Cedex 1, France
☎ (1) 820.61.64, Telex 610789 ASFRA

United States of America
Butler Cox & Partners Limited
216 Cooper Center, Pennsauken, New Jersey 08109, USA
☎ (609) 665 3210

Switzerland and Germany
Butler Cox & Partners Limited
Morley House, 26-30 Holborn Viaduct, London EC1A 2BP
☎ (London) 583 9381

Italy
Sisdoconsult
20123 Milano – Via Caradosso 7 – Italy
☎ 86.53.55 / 87.62.27, Telex 311250 PPF MI

The Nordic Region
Statskonsult
PO Box 4040, S-171 04 Solna, Sweden,
☎ 08-730 03 00, Telex 127 54 SINTAB