# INFORMATION RESOURCE/ DATA DICTIONARY SYSTEMS

Henry C. Lefkovits
Edgar H. Sibley
Sandra L. Lefkovits

# INFORMATION RESOURCE/ DATA DICTIONARY SYSTEMS

Henry C. Lefkovits
Edgar H. Sibley
Sandra L. Lefkovits

QED Information Sciences, Inc.
Wellesley, Massachusetts 02181

INFORMATION RESOURCE/DATA DICTIONARY SYSTEMS



Printed in the United States of America

Library of Congress Catalog Card Number: 83-60770

International Standard Book Number: 0-89435-068-4

# PREFACE

Dictionary systems have the potential of being a powerful tool for the control and management of data and information resources, as well as in the planning and development of information systems. Since the number of new dictionary projects is continually increasing, we must conclude that the problems that the use of a dictionary system addresses are very real, and that the need to solve these problems is a pressing one. We are also seeing today the rise of a number of new technologies that have a severe impact on the manner in which the business of building and using information systems is handled, and the dictionary system is beginning to assume a central controlling and coordinating role in these new architectures.

A reflection on the importance that is being attached to the role of dictionary systems is the fact that two parallel efforts at specifying a standard system are currently in progress. The American National Standards Committee has set up a Technical Committee (X3H4) to develop such a standard, and the National Bureau of Standards is pursuing the development of a Federal Information Processing Standard for Data Dictionary Systems; this latter work is being carried out by Alpha Omega Group, Inc., with two of the authors of this report being deeply involved in the effort.

Experience forces us to admit that there exists a large number of projects where the use of a dictionary system has fallen far short of expectations. The causes for these failures fall into two major categories: the lack of proper planning for the project, and/or an unrealistic assessment of the benefits and costs. The work that is presented here attempts to provide a better understanding of dictionary systems, the functional roles that they can support, their features, and the characteristics of a number of leading products currently in the marketplace.

The first chapter attempts to provide the common thread that exists between the problems that can be addressed through the use of a dictionary system and the characteristics of a planning methodology that will significantly improve the chances of a successful project. The second chapter introduces the components of a dictionary system, presents the terminology that is used throughout this book, and describes the facilities and features that are commonly found in a dictionary system. This is followed by chapters, each one of which contains the description of a commercially available system. The format we have used in these descriptions generally follows the outline of features discussed in Chapter 2, but varies to some extent from one system to another. We have tried to present a valid description of the features and facilities of each system, and have used the level of detail appropriate to the individual system to achieve this goal. In no case should the description be thought of as replacing the vendor supplied documentation of the system. In most instances, a detailed description of the syntax of

the command language has been omitted. There is no significance to the order in which these systems are presented.

We are using a consistent terminology (the one developed in Chapter 2) in the description of all of these systems; this terminology is sometimes substantially at variance with that used in the documentation of a system. We also use "database" as a preferred term whenever we are not referring to a "data base" of a specific DBMS where that spelling is used.

The material presented in Chapters 3 and 8 is held to be proprietary by the companies marketing these products. As part of the agreement allowing the use of this material in this work, these chapters have been reviewed by the companies marketing the respective products and have been found not to contain errors of fact. Other than the material in these chapters, the findings reported here are solely the opinion of the authors and are based on the evaluation of existing material.

The authors would like to express their appreciation to all the companies whose products are described here for their cooperation with this effort. In particular, we would like to acknowledge the efforts of Barbara Hobbins of IBM, Lawrence Tindell of MSP, Inc., Jack Grady and David Thole of Cullinet Software, Inc., Jim Watts of ADR, Cheryl Traver of Intel, George Gajnak of Cincom Systems, and Ben MacCarley of UCC, in the review of a draft of the chapter dealing with the system of their respective company. Special acknowledgements are due to Stuart Douglas Third and Caroline Margaret Johns of MSP, Inc., and Roger Bruylands of ADR, for their work in suppyling us with special material used in these two chapters. We also want to express our appreciation to Dr. Anthony J. Winkler of Alpha Omega Group, Inc. for his critical review of some of the chapters, to Edwin F. Kerr of QED Information Sciences, Inc. for his efforts in the coordination of the production of this work and to Donna St. Martin of QED for her patience and conscientiousness in completing the art work.

Henry C. Lefkovits
Alpha Omega Group, Inc.
Harvard, Massachusetts

February 1983

# CONTENTS

# 1
# DICTIONARY SYSTEMS — A MANAGEMENT PERSPECTIVE

CONTENTS

# DICTIONARY SYSTEMS — A MANAGEMENT PERSPECTIVE

## 1.1 INTRODUCTION

The concept of a "dictionary system" has existed in the data processing industry for a number of years. Use of such a system consists, basically, of an attempt to capture and store in a central location, definitions of data and other entities of interest. The rationale for such a system has always been that:

- It contributes to better control of data as it is created, stored, and used;
- It assures better documentation; and
- It has a very positive effect on the quality of systems that are built, both in terms of user satisfaction about the functionality of the system and its maintainability.

The spectrum of applications for which dictionary systems have been used has always been wide and has widened substantially in the past few years. On one end of this spectrum has been what is commonly referred to as a "data element dictionary". Here the scope of interest has been to capture and document data elements, their definitions, and some of their descriptive attributes. Optionally, this effort would also include some logical groupings of data elements. Such an effort, which seems to be a small subset of the type of scope which we will discuss later, can nonetheless be tremendously productive, as well as cost-effective. Consider the process of gathering requirements for a new system which is to be built; such a data element dictionary provides the vocabulary that can be used between the systems analyst and the end-user. Adherence to a commonly understood and generally agreed upon terminology will contribute substantially to assuring that the statement of requirements reflects the true needs of the end-user and the enterprise. Equally, for data elements which are shared by different users and perhaps different organizational components, availability of commonly agreed upon definitions of data elements will help in clearing up misunderstanding of terminology or, even more importantly, may prevent such misunderstandings from occurring.

The next level in the spectrum of usage of dictionary systems can probably be traced to two separate, though related, trends. One of these is that, once a data element dictionary is available, it may be extended to include information as to how and by whom the data elements are used. This introduces the necessity to use the dictionary as a repository not only of data element definitions but also for:

- The definitions of other data constructs, such as records and files;

- The definitions of processes, such as programs, or perhaps manual processes; and

- The definitions of users of data, whether they be individuals or organizational components.

Along with the definitions of such other entities, the dictionary also began to be used to contain the cross-references between them to record their usage, and perhaps organizational responsibilities.

The second trend that contributed to this extended usage of a dictionary system was the gradual migration away from the use of conventional files toward the concept of a central, integrated database under the control of a Database Management System (DBMS). In the use of conventional files, each application was seen as owning its data. There was some understanding of a need to document and manage the contents of these files, but the effectiveness of this was generally minimal; control across the individual applications areas was lacking and data duplication and its associated maintenance caused problems. In the context of a central database, the problem of duplication of data (and inclusion of unnecessary data in the database) was generally also a problem, though this was not always obvious in large organizations. A central control mechanism therefore became essential, and a dictionary system is an effective tool to aid in the solution of this problem.

At this point, it is clear that a dictionary system is helpful in the environment of a DBMS, even though it can probably be argued that this utility is not particularly greater than in the conventional file environment. But there is more to what has happened.

One of the rationales for a DBMS is that it will provide control over data and effective management of this data. This is particularly important in an environment where this data is to be shared not only by different systems, but also by a wide range of users. A basic concept of a DBMS is, then, to provide a centrally located set of definitions of the data that is to be shared in order to assure that different users will access this common data with a set of consistent definitions. In the majority of the DBMS's two problems are encountered with this concept:

1. The central definitions of the data, to a large extent, only contain information which is required by the DBMS itself in carrying out the functions it performs. As such, it contains much data that is required by the DBMS but very little that is of interest to the user. But in many cases it is precisely those aspects that are important to guarantee consistency in the use of data, and a dictionary system is required to supply those facilities. An example of DBMS required data is in the area of descriptions of these data elements because they are required in the memory management. Most

of these characteristics are, however, not of great interest or use to people: their communication on shared data involves the development of an understanding of the meaning of the data elements. Equally important is the fact that the same data element is often referred to by different people by different names (the synonym problem). Conversely, the same name is sometimes used by different people to mean different data elements (the homonym problem). The dictionary system, working in what may be thought of as an adjunct mode to the DBMS, is able to alleviate this shortcoming in the DBMS.

2. The existence of the central definitions in the DBMS gives rise to another problem. In the vast majority of DBMS's there is no provision for the management of the central definitions. It then follows that the actual control over the data is substantially weakened since there is a lack of control over the definitions of the data. It is here again that a dictionary system, when operating in conjunction with a DBMS, strengthens the true capabilities of the DBMS.

It can therefore be seen that both of these trends have influenced the expansion of the scope of usage of the dictionary system.

A further quite independent development has also influenced the usage of dictionary systems. Without doubt, the data processing field has attained a greater degree of maturity. One effect of this maturity is manifested in a greater emphasis on control and cost-effectiveness in product development as well as in an operational sense. In days past, the typical "data processing department" was a group of people, technically oriented, who were adept in converting manual processes into computerized systems and procedures; this department was merely a "translator" into computer programs from what had existed before. There also existed an aura of what could almost be perceived as an element of "magic", which meant that normal policies and procedures of the enterprise somehow did not quite apply to the data processing department. This was particularly true when it came to its management. Since the basic interest of the people in the department was technical, much too little attention was paid to the management of the department, and to the control-oriented aspects of the products that were produced.

This situation became worse when it was realized that a simple translation from manual processes was not enough. These processes and their design had been influenced by the very fact of their being manual, which imposed strong limitations on their functionality. Once these limitations were removed by the use of automated machinery, it could be seen that entirely new procedures were required by the users of the data, and that much more attention had to be paid to the task of designing systems.

 

Clearly, this was a situation that could not last. A measure of maturity occurred when better management practices were implemented and the organization realized that it was important to control both the development process of systems and the usage of data that was a part of these systems. One of the major developments was the adherence, at least in some fashion or another, to the concept of a System Life Cycle, which divides the development process into a set of well-defined stages. It is then possible to define the activities that must take place in, say, the Requirements Definition stage or the System Design stage and to introduce the concept of reviews that take place, at least, at the end of each stage. A successful review would then be necessary before proceeding to the next stage. The increasing use of a dictionary system must be interpreted as another aspect of this degree of maturity, even though it must be made clear that we still have a long way to go on this subject and that the use of dictionary systems, at least on the average, is still very far from their potential. Later, in Sections 1.5 and 1.6, we will also discuss the use of a dictionary system as an aid to managing the System Life Cycle.

We are now entering a new era, where it is being recognized that the problems yet to be solved are really even larger than those we have been addressing, and that the scope of information processing extends substantially beyond the data processing department. Many factors contribute to this recognition; among the more important are:

- The greater involvement of the end user in most phases of the information processing process.
- Economic pressures on most enterprises that require increased productivity at all levels.
- Efforts aimed at "automating" the office, which introduce concepts of information processing that are substantially different from traditional data processing.
- The advent of the mini- and micro-computer, the resultant distribution of data and processes, and the greater reliance on communications links.
- The availability of advanced technology, in both hardware and software, offering many new opportunities for both success and failure.

The methodology that attempts to solve this expanded set of problems in an integrated and coordinated manner is generally referred to as **Information Resource Management (IRM)**. In many ways, this is still a "buzz word", but we believe that we are seeing the beginning of a profound and significant change in the way that information systems will be viewed in the eighties.

It is equally clear that a dictionary system will play an important role in this process, and hence the title of this book has been chosen to read Information Resource/Data Dictionary Systems. The American National Standards Committee has delegated to a Technical Committee the responsibility for defining a standard for an "Information Resource Dictionary System", which is to be a dictionary system that contains features believed to be tools required for Information Resource Management. In the remainder of this book we will simply refer to all such systems as "dictionary systems", as the transition from a Data Dictionary System to an Information Resource Dictionary System is gradual and influenced by the uses to which the system is put.

We will return to a discussion of Information Resource Management in Section 1.3 of this chapter.

## 1.2 INFORMATION AND DATA AS AN ENTERPRISE ASSET

It is a well-accepted practice in most enterprises to have numerous and explicit policies to guard, protect, control, and manage both tangible and intangible assets of the enterprise, be they financial assets, materials, facilities, or personnel. These policies, and the responsibilities and authorities assigned in them, are normally reflected in the organizational structure of the enterprise in that major categories of assets correspond to major organizational components.

Even though it may generally be difficult to quantify it, data and information are also a valuable resource of the enterprise: money is spent in collecting or generating data and information; they are important for use in the control of the enterprise and in decision-making; and the knowledge and availability of data may be an integral part of the operations of the enterprise. In this sense, the proper management of the information resource is vital to the success of the enterprise.

Our usage of the terms "information" and "data" complies with modern standard practice. Information relates to a human process involving the understanding of facts. On the other hand, these facts may be recorded on some medium, such as paper or magnetic tape, and what is recorded will be termed data.

There is little question that the enterprise data base represents a substantial investment. It is difficult to assess the costs associated with the existence of data in the normal course of business, perhaps just as difficult as it would be to assess the costs associated with the absence of this data. Certainly, however, should a detailed study be made of the costs of collection, maintenance, and storage of this data (which would have to include the people, machines, processes, physical assets, and procedures involved), there is no question that these costs would far exceed estimates that one might collect superficially. In view of the costs associated with maintaining the data base of the

 

enterprise, and the tremendous penalty that could be incurred in case it is not available, it must be asked why many enterprises do not exert greater efforts to control and manage this asset.

If we were to look at procedures that exist for management of the data and information resource within data processing organizations, we would find that some installations have installed sophisticated security measures to protect the data processing installation, but this protection tends to be largely aimed at the physical installation. Largely due to some well publicized frauds, more and more installations dealing with financial data are installing access control measures which are aimed at preventing unauthorized usage and alteration of the data. Any of the above is, however, a far cry from the kind of procedures that are in place for the management of other assets. It is probably a safe bet that in most installations there exists a better inventory of the chairs in which people dealing with data sit, than an inventory of the data with which they deal. Equally, there is probably a better description available of where the chairs are located than where the data can be found.

The above example may appear to be humorous, but it is representative of many analogous situations. Consider the following:

> When an organization maintains a stock of a certain spare part of a certain make and year of truck, it can be assumed with reasonable certainty that it uses that kind of truck, and probably the stock and reorder level have been determined with the aid of some formula on the basis of number of trucks and failure rate.
>
> Consider the case of some data element: is anyone in the enterprise using it? We may do some research, and find out that it appears in a certain report. Who gets the report? This may take some more research, and it may be the manager of a certain department. Is it possible that it was ordered to be produced by an incumbent in this position several years ago (who possibly was working on some theory that was not passed on to the successor), and that a series of successors (perhaps fearing to exhibit ignorance) simply threw away the reports that were produced with unerring regularity?

Perhaps the example is too absurd, and assuming that none of this is true, is there positive knowledge that collecting this data is needed for the operations of the enterprise? Is there a policy that can be applied to determine the frequency with which this data must be backed up or how many archival generations are to be retained? Is it just possible that this data is collected in two (or even more) locations, but where each location knows it by a different name? The simple point that we want to make is that this would never happen with chairs or spare parts for trucks.

 

Many similar questions can be asked when looking at the documentation for programs versus the documentation of data in an enterprise. Many installations have elaborate standards for program documentation, which may or may not be observed, but seldom is a fraction of that concern expressed for data. Documentation of data often exists in a fragmented and incomplete form in data definitions within a program, or even imbedded in the program logic.

There are no simple ways of answering the foregoing questions. However, the almost explosive speed with which the entire data processing field has grown can certainly show the value of information systems and their associated data. Often it appeared that nothing ever stood still and unchanged long enough to be evaluated. The advent and increased use of DBMS's, along with the concept of central definitions of databases which are shared, has ameliorated the problem to some extent, but as we discussed in the preceding section, this did not solve some problems, and has also created other problems.

The enterprise's data base must be recognized as an asset to be controlled and managed, and a method of accountability for this asset must be recognized. This recognition, although not solving the problem of "how", is a necessary first step.

The very same concerns that we have voiced in this section about data can also be applied to other entities that play a role in the information systems of the enterprise, and this will be discussed in the next section.

Later we will discuss the role of dictionary systems in attempting to find solutions to these problems. We must, however, point out that a dictionary system is a tool, perhaps the most powerful one available today, that can aid in the solution. But a dictionary system is a tool, not a solution. If properly used it can be an effective ally, but if improperly used it can negate the very goals for which it was created.

## 1.3 INFORMATION RESOURCE MANAGEMENT

Before we discuss Information Resource Mangement in this section, some clarification is in order. Our main target is dictionary systems, and whereas it may be intuitively clear that these two subjects are related in some fashion, the nature of this relationship is probably not clear at all. We expect to show that a dictionary system project is in fact an Information Resource Mangement project; placing the dictionary system in this context will give it more meaning. Equally, the real issues that must be faced in a dictionary system project will become highlighted.

In the previous section we have discussed the value of data and information, and have argued for the need to establish policies and procedures to manage these resources. We have largely limited our discussion to data in relation to computer systems. Here we will