Data dictionary—More on the impossible dream

by MICHAEL EHRENBERGER
Cincom Systems, Inc.
Cincinnati, Ohio

ABSTRACT

This paper discusses the role of Data Dictionary software in the computing function. It specifically discusses the evolutionary process which brings about the need for a Data Dictionary. It goes on to describe the major components and requirements of Data Dictionary software along with its interaction with data base and the data base administration function.

The advantages, benefits, and potential drawbacks from the misuse of this software are also examined.

INTRODUCTION

Since the initiation of computing, the data processing industry has been preoccupied with hardware selection, programming techniques, project management, and the like. Computing, like other emerging disciplines, tended to orient its management philosophies around the physical hardware rather than the function the hardware performs. Recently, this concept of computer management has taken on a new meaning. The computing industry is no longer preoccupied with computing; rather, it is concerned with the management of a vital corporate resource—data. For that reason, whenever we speak about management in the context of data processing or computing, it’s important that we realize that data is the thing that we’re managing.

This paper will deal with the acceptance of a new tool used to control that data—The Data Dictionary. At the same time, it will try to address managerial techniques whereby we can more effectively manage the data within our own department.

Over the last four to five years the whole concept of data base and data communication systems has emerged. This particular philosophy has been adopted by a majority of the data processing installations throughout the United States. Because of the tremendous acceptance of data base and data communication systems, the amount of processing required of a computer has expanded geometrically. At the same time, data, standards, procedures, programs, systems, reports, and personnel within data processing operations have also expanded at a geometric rate. The management of the computer operation itself is an increasingly challenging job and one which calls for new and different skills. A new position has been established to deal with the management of data—that of the data base administrator. A key tool of a data base administrator is the Data Dictionary. It is the intent of this paper to try to deal with the role of the Data Dictionary in the management of the computing function. In order to achieve that, the following topics will be covered:

1. The evolution of the need for Data Dictionary
2. Management needs to be addressed by the Data Dictionary
3. The advantages and benefits of the Data Dictionary

THE EVOLUTION PROCESS

Computing and data processing is not unlike the growth of other disciplines in business today. You might reflect back on the emergence of such disciplines as production control, manufacturing management, and the like. It was only after the intensive studies of people like Frederick W. Taylor and Henri Fayol that manufacturing management gained wide acceptance throughout industry in the U.S. At the same time, it was not until the introduction of double entry accounting systems that consistent accounting controls were established. Computing and data processing, like other management disciplines, is now recognized as an important part of the organization. This recognition has not been sudden; rather, there have been various stages of development. These stages have been described by Dr. Richard Nolan of Harvard University. The question of “when Data Dictionary” can be answered by determining what stage of development a particular data processing operation has achieved.

According to Dr. Nolan, there are four, possibly five, development stages. The first stage of growth is the initiation stage when we first acquired the computing power within our organization. During this initial stage, the applications are primarily oriented towards accounting; the personnel we hire are oriented towards the effective use of a particular piece of hardware; and the management itself,
typically, is management contained within the functional area which approved that piece of computing power. During this initiation stage, the computer is an under-utilized piece of capital equipment, probably only used 50 percent of the time. This under-utilization provides the rationale for entry into the second stage. During the second stage, we begin to expand. Applications are proliferated in all areas; the personnel within data processing become highly specialized in programming languages. With respect to management, we are oriented towards selling computing services. It is obvious at this stage that we are managing a piece of hardware—the computer itself. There is also a dramatic rise in the budget associated with data processing. This increasing cost led us to the third stage.

The third stage is the stage of control. Typically, no new applications are done; rather, existing applications are rewritten in a native mode or to be more efficient. From a personnel point of view, this is the age of the operating systems. Emphasis is placed on systems programming and the fine tuning of hardware/programs. From a managerial point of view, there is tremendous upheaval, i.e., reorganizations, centralization, de-centralization, etc. It is also during the stage of control that people recognize the vital importance of data processing within an organization. It is at this point that users begin to ask data processing to modify existing systems, to enhance systems, etc. Conversion is fast becoming a way of life as users recognize the importance of data. The integration of data, systems, and programs brings about the fourth stage and the need for new, sophisticated tools. It is at this point in time that some new concepts are introduced, i.e., the concepts of data base/data communications. Along with these concepts comes a dramatic change in the role of computing within the organization. Management of the computing function is now a key job in the organization. This is certainly true when you take a look at the effect that this particular function can have on the overall profit and loss of a corporation. Companies who are cognizant of this are well into the fourth stage. The fourth stage is characterized by a recognition that the responsibility of data processing managers is the management of data, not management of the computer. It is not unlike the basic recognition of the production manager that his role is not the management of machines; rather, it's the overall management of production.

THE NEED FOR A DATA DICTIONARY

The Data Dictionary is designed to manage the data within the data processing department. Like the story of the cobbler who made shoes for his family only after he had satisfied the needs of the villagers, the Data Dictionary provides to data processing the same needs that data processing has historically provided to user departments, i.e., the management of their data.

The onset of integrated data management brought with it a more complex environment and an environment which required new and different tools. Data Base provides functional integration while the Data Dictionary provides the control of that Data Base. Data Dictionary like Data Base is a prerequisite for evolution into the fourth stage. The Data Dictionary should contain information about, and the relationships among the entities within the realm of data processing. These include:

- Data Files
- Reports
- Departments
- Personnel
- Data Fields
- Systems
- Projects
- Standards
- Programs
- Users
- Transactions
- Source Documents
- Data Bases
- Security Levels

The Data Dictionary should provide utilities to automatically generate information about the above. Utilities which scan existing programs or libraries could be used to generate a good deal of the required information. Other information will need to be researched and/or established with human intervention.

The Data Dictionary should also play an active role in the day to day operations. For that reason, a Data Dictionary should include the following features:

Automatic program set-up

A programmer should not be concerned with coding a "data division" or "I/O" areas within a program. This information should be stored within the Dictionary and automatically copied or invoked at compile time. This feature substantially reduces programming and maintenance time while insuring security and control over the data available to programs.

DBMS interface to the data definition language

Data Base Management Systems and Data Dictionaries must fit hand in glove. For that reason, the Data Definition required for each data record, data set, and data base should be stored within the Data Dictionary and automatically generated on request. Once again, this feature is a necessity for effective data base administration.

On-line data dictionary access

The trend toward cardless systems requires that the Dictionary provide on-line editing and update. Currently on-line editing and validation is built into each on-line program. Access to the Dictionary would all but eliminate this redundant programming effort. The savings are obvious. On-line updates of the Dictionary allow dynamic changes to editing criteria, security levels, passwords, etc. With the advent of tighter privacy legislation, this capability will be required.

Automatic report generation

The Data Base Administrator will need various reports about the Dictionary. Current information and relationships about systems, applications, fields, records, files, pro-
grams, users, terminals, etc., are required. The report feature should be an embedded part of the Dictionary. In addition to the standard reports, the Dictionary should be constructed such that non-standard, ad hoc reporting is possible.

Data base documentation support

This rather nebulous sounding characteristic may well be the most important. In essence, it allows a DBA to ensure data base usage and programming standards. This is accomplished through a comprehensive description of a company's data base. It encompasses the following components:

- Complete Attribute Description of All Entities.
- Automatic Relationships Between Entities.
- Security support
- Entity security applies to the protection of data files, fields, programs, etc. Data Dictionary security applies to the protection of 'the Data Dictionary. It is important to note that we be able to accurately determine and quantify the impact of user-requested change. A Data Dictionary, as previously defined, provides answers to these questions. Interestingly enough, the ability to accurately assess the impact of change is a key indication of the level of development of any management discipline.

EDP audits are always in vogue. Consistent documentation and standards are an important aspect of such audits. The Data Dictionary eliminates much of the manual labor associated with documentation while substantially upgrading the quality of the documentation. The cataloging of entity attributes provides programming standards, naming standards, data base standards, security, and integrity. It also provides a common repository of data about the data base thus allowing tight control by a Data Base Administrator.

Possibly, the one drawback of the Data Dictionary is in the overhead associated with day-to-day operations. If all access to data requires an additional access to the Dictionary, then the overhead could be substantial. Like many new concepts, the Data Dictionary should be used with enthusiasm but with a measure of discretion. As new technologies emerge, the overhead of accessing data will probably reduce to core-to-core transfers. At that point, the full concept of data base—data dictionary—can and will be a reality.

In conclusion, it is my opinion that data processing had indeed emerged through four stages. In fact, Dr. Nolan has now published a new article indicating the fifth stage of development. The fourth and fifth stages are obviously dependent upon the tools required for effective management of those stages. Data base/data communications are keys to the success and transition through stages four and five. The Data Dictionary is a key ingredient to this success. From all indications, the major corporations in the U.S. will be adopting the concepts of Data Dictionary in the next year. In our opinion, it is a key ingredient to managerial success in the computing function.

REFERENCES
