

The Role of the Data Dictionary

by George Schussel

Their primary purpose is to aid in the control of the corporate data resource, but they also help in reducing programmer error and even in documentation.

For several years, not only has it been in vogue but also considered good business practice to devote closer attention to the management of data as part of the corporate dp function. Witness to this is the rapid emergence and acceptance of data base management systems (DBMS) as primary control software in manipulating information files. Many feel that the arrival and acceptance of data base concepts and data base management systems is the single most important happening in the data processing field since the development of operating systems.

Data dictionaries are part of this "happening." Altogether different types of software products from DBMS's, these dictionaries are now becoming widely available and readily accepted as primary tools for better data management. They can be used with or without DBMS's too, as the two package types are complementary, not mutually exclusive.

A data dictionary is a repository of information about the definition, structure, and usage of data. It does not contain the actual data itself. Simply stated, the data dictionary contains the name of each data type (element), its definition (size and type), where and how it's used, and its relationship to other data.

The purpose of such a tool is to permit better documentation, control,

and management of the corporate data resource, goals which may or may not be achieved through the use of a DBMS. Advanced users of data dictionaries have found them also to be valuable tools in the exercise of project management and systems design.

The rationale for data dictionaries is different from that for the formation of a formal data base and a DBMS, which is, of course, to achieve data independence. Historically, DBMS's have gone a long way toward allowing independent management of actual data occurrences and of the programs that manipulate and access such data. This independence of control has resulted in substantially enhanced usefulness of the data.

Information need no longer be stored in fixed master files which are suitable for only one application. Through a DBMS' ability to take one physical description of data and multiply it into a large number of alternative, logical descriptions, it's possible for the same data to serve a multitude of users, each having a slightly different picture of the data. Used in this context, data has not become locked away in the programs but has become more of a corporate resource.

That is what happens when a DBMS is properly used. Frequently, however, they have been used for their storage and retrieval capability primarily, as

alternatives to ISAM or VSAM, for example. This is not a recommended use of DBMS's but has happened frequently, nonetheless. Unfortunately, the mere presence of a DBMS does not mean that data are controlled or managed any better than in second generation files; this enhanced control can only be brought about by edp policies directed toward improved data quality.

What are they good for?

It is here that a data dictionary comes in to play. Containing all of the definitions of the data, the dictionary becomes the information repository for data's attributes, their characteristics, their sources, dispositions of usage, and interrelationships with other data. It serves to answer questions like:

1. What kind of validity tests have been applied to this data type?
2. Who is authorized to update it?
3. What modules, programs, and systems use this data type?
4. What are the valid ranges of values for this data?
5. What security level is applied?
6. Who is allowed to access the data?
7. By what other names is the data type known in various application environments?
8. In what reports does this data type appear?

DATA DICTIONARY

9. What is the input source for this data type?

Frequently we find data dictionaries being used in combination with DBMS's. As suggested earlier, this is not necessary and, in fact, enlightened shops in the 1960s that did not have a DBMS installed were already using data dictionaries. Also, today a majority of the estimated 4,000-plus North American users of DBMS's do *not* yet have data dictionaries installed, although it is my guess that most of these sites will eventually migrate to become dictionary users.

When used in combination with a DBMS, a dictionary can become exceedingly powerful with automatic interfaces providing:

1. The ability to *automatically* copy definitions and generate entries from existing information contained in COBOL and PL/1 programs and/or the DBMS schema and subschemas. Used in this way, naturally, the dictionary can only perform clerically and will not streamline an already existing mess of incongruent definitions.
2. Alternatively, these definitions may exist in the dictionary and we can *automatically* generate schema and subschema sections from the entries. But not all dictionary/DBMS interfaces work in both directions. So it's important to analyze each particular interface to verify what capabilities are supplied; maximum value can be achieved only through this two-way interface.

Clearly, while the use of a DBMS to manage the data base does not by itself generate any documentation improvements, the use of a dictionary provides documentation of a quality and form that is simply not available through less formalized procedures in the dp environment.

Also, the availability of an accurate and widely used dictionary will allow project managers to improve new project estimates. This applies especially to maintenance projects where experience has shown estimates to be off frequently by as much as 100% to 300%. Through the use of the dictionary, the project manager can learn how often a data type is used, what program modules use it, and how it is presented to those modules. In a maintenance environment, this sort of quickly available information can be invaluable in calculating the impact and cost/benefit of

TYPICAL DATA DICTIONARY PRODUCTS

Vendor	Dictionary Product	DBMS Interfaces	DBMS Required	Access Security	On-Line Access	Dictionary Entry Generated Automatically?	Definitions From Dictionary Automatically?	Approximate Pricing (for standalone)
MRI Systems	Control 2000	System 2000	System 2000 (must license)	yes	TP 2000	promised	promised for COBOL PL/1 FORTRAN BAL	\$15,000
Synergetics	Data Catalogue	IMS TOTAL	no	yes	TSO (inquiry only) CICS	yes	COBOL PL/1 BAL	\$12,900-\$18,900
Cincom	Data Dictionary	TOTAL	TOTAL	yes	no	no	no	\$11,000
MSP, Inc.	Datamanager	IMS TOTAL IDMS ADABAS	no	yes	Intercomm Taskmaster ROSCOE IMS/DC CICS TSO CIMS ETSS	yes	COBOL PL/1 BAL MARK IV	\$9,900-\$24,900
IBM	DB/DC Dictionary	IMS DOS DL/1	IMS or DOS DL/1	no	IMS/DC (VS)	yes	COBOL PL/1	\$290-\$580/mo
Cullinane	IDMS Dictionary	IDMS	IDMS (no license fee)	no	TSO IDMS Query	Sept. 1977	COBOL PL/1	\$15,000
Arthur Anderson	Lexicon	IMS TOTAL IDMS	no	no	TSO	no	COBOL PL/1	\$10,000*
Univ. Computing	UCC 10	IMS	IMS	no	IMS/DC	no	COBOL	\$18,000

Table 1. *This software is installed only in conjunction with Arthur Anderson & Co. services.

DATA DICTIONARY

any proposed change.

Programmers learn to appreciate the advantage of the tool by automatically generating file descriptions through COPY's into their programs. This is because the COPY facility reduces the total amount of coding that is required and, of course, by COPYING rather than re-entering, errors are reduced.

Who makes them?

There are a number of important differences in the eight packages outlined in Table 1, though all are marketed as general-purpose data dictionaries. Cincom Systems' Data Dictionary was introduced in 1976 and has been marketed primarily to existing customers of Cincom's TOTAL data base system. Control 2000, likewise, has been offered by MRI Systems Corp. for use with its System 2000; it was announced in 1977.

The Data Catalogue (Synergetics, 1973) and ucc Ten (University Computing, 1970) are marketed to users of IBM's IMS (and in Data Catalogue's case, also to users of TOTAL).

DATAMANAGER was introduced in 1975 by MSP, the only vendor in Table 1 whose principal product is a data dictionary. DATAMANAGER has been interfaced with most of the data base management systems used on IBM's mainframes.

DB/DC Dictionary from IBM and IDMS Dictionary from Cullinane are both fairly recently announced products (December 1976), and DB/DC is intended only for IBM's proprietary DBMS's (IMS and DOS DL/1). Cullinane's product, although based on the proprietary IDMS, IDMS Query, and CULPRIT packagers, doesn't require licenses for those packages for dictionary use only.

The Lexicon Dictionary is part of a powerful software package which includes generator capability, query capability, plus input processor, data extractor, and table maintenance systems facilities. Arthur Anderson and Co. the vendor of Lexicon, is a major accounting firm substantially different from the other vendors who are all primarily concerned with computer hardware or software. The firm has not been aggressive in selling Lexicon outside its own customer base of audit clients, usually supplying it as part of a combined package including consulting services.

The prices of these packages tend to be relatively conservative compared to the prices of DBMS systems. Perpetual paid-up license fees run in the range of \$10,000 to \$25,000 on a one-time fee basis, with annual maintenance charges running around 10% of the original license amount.

What should they have?

In looking for a dictionary package, questions to ask include:

1. *Is it based on a specific data base management system?*

One important thing to look for is whether the product includes, as part of its own file structure, the use of an already existing DBMS. This is important in ascertaining the usefulness of the dictionary. A true network-based DBMS internal structure may offer more flexibility than the more conventional ISAM or VSAM-based dictionaries, but at an increased overhead cost.

Also, for those dictionaries that do use a DBMS as a file controller, it may mean that rights to use of the DBMS must be acquired before the dictionary may be run—not necessarily a desirable feature (and not true of Cullinane's product). Then too, those based upon DBMS's are usually proprietary products marketed by the same vendor that generated the DBMS, and intended for an environment which consists primarily of an interface to files controlled under the DBMS.

A review of the packages on the market shows that generally (but not always) the packages which support their own generalized file structure rather than being based on some other have been interfaced to a wider variety of DBMS's and do a better job in supporting standalone individual files.

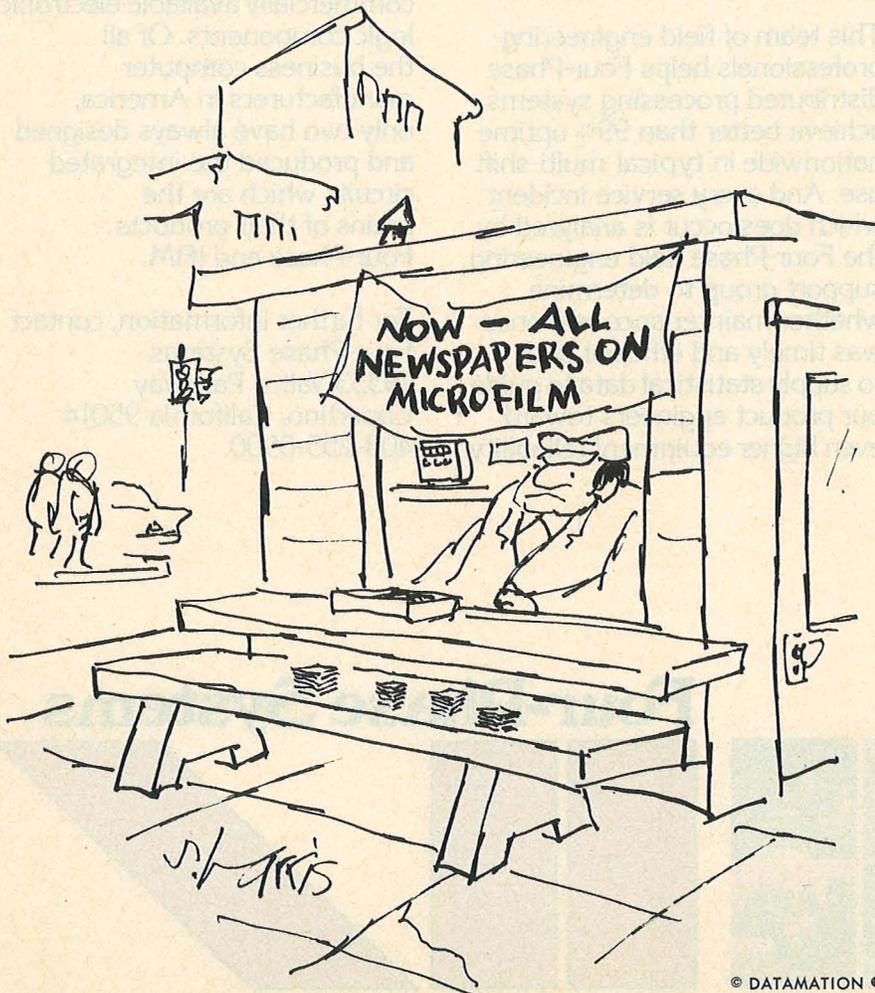
2. *Has it been interfaced to a wide variety of DBMS's?*

With the advent of the large scale disc drives and the wide availability of DBMS software, it is clear that the majority of large scale users of data will be moving to (if, in fact they haven't already) the use of DBMS (or often several different ones for mature users). Accordingly, a data dictionary can only be significantly useful if it is automatically interfaced to all of the DBMS's being used.

Also, importantly, when these interfaces don't exist, the potential exists for updating information in the DBMS while forgetting to make the appropriate update in the dictionary, or vice versa. Once the two systems get out of phase, there will be a tremendous loss of credibility, in addition to other severe problems.

3. *Is it easy to use?*

A number of non-dp personnel may use the data dictionary. James Martin (*Principles of Data-Base Management*,



© DATAMATION •

DATA DICTIONARY

Prentice Hall, 1976) has postulated that active users of a dictionary should include general management, auditors, spontaneous users of terminals, application planners, data base management, the data administrator, systems analysts, and programmers.

Clearly it is difficult to imagine an environment where, at a minimum, data administrators, systems analysts,

DATA DICTIONARY VENDOR INDEX

For more information about the data dictionaries mentioned in this article, either contact the vendors listed below, or circle the appropriate number on the reader service card bound into this issue.

Arthur Anderson & Co.
69 W. Washington Street
Chicago, IL 60602
(312) 346-6262
CIRCLE 300 ON READER CARD

Cincom Systems, Inc.
2300 Montana Avenue
Cincinnati, OH 45211
(513) 662-2300
CIRCLE 301 ON READER CARD

Cullinane Corp.
20 William Street
Wellesley, MA 02181
(617) 237-6601
CIRCLE 302 ON READER CARD

IBM Corp.
1133 Westchester Avenue
White Plains, NY 10604
(914) 696-1900
CIRCLE 303 ON READER CARD

MRI Systems Corp.
12675 Research Blvd.
Austin, TX 78759
(512) 258-5171
CIRCLE 304 ON READER CARD

MSP, Inc.
594 Marrett Road
Lexington, MA 02173
(617) 861-6130
CIRCLE 305 ON READER CARD

Synergetics Corp.
One DeAngelo Drive
Bedford, MA 01730
(617) 275-0250
CIRCLE 306 ON READER CARD

University Computing Co.
8303 Elmbrook Street
Dallas, TX 75247
(214) 688-7100
CIRCLE 307 ON READER CARD

programmers, and auditors would not all independently wish to access information on the dictionary. In this environment, a free-form or easy to use English-like query language in an on-line capability can be very important.

4. *Does it support on-line interaction?*

Some dictionary systems have been designed to run only in a batch environment, although some have been interfaced with one or another teleprocessing product for availability in an on-line environment. Unless management wishes to print out the dictionary's contents on a periodic basis, generating a lot of paper, on-line access is key. (When the dictionary is printed you're never sure the printed version you have is the latest. Only on-line inquiry can verify this.)

Also, a facility to support "keyword" query is desirable. The other part of on-line interaction is the ability to update; some systems may support query but not on-line update.

5. *What kind of reports are available?*

All of the packages come with many different standardized report formats. These are useful, but some of the packages also have free-form reporting capability with the equivalent of RPG report generation. Also, a USER EXIT capability for applying other RPG packages to a dictionary file is sometimes available. These additional abilities are essential for custom tailoring reports.

6. *What about automatic interface to existing programs?*

A dictionary package will usually be introduced in an environment where a large number of COBOL or PL/1 programs already exist. The ability to automatically create data entries from the data divisions of these programs is a significant aid.

7. *What kind of security is available?*

Some feel that because the dictionary gives a model of the corporate data, that the perfect fraud could best be tested by analyzing it through the relationships contained in the dictionary. In any case, with all information concerning the corporate data resource located in the dictionary data base, it is clear that responsible management requires control of access to those definitions. Note that the level of security mechanisms offered by the various packages differs substantially.

8. *"What will it do to the overall systems if we change the structure of this data element?" answers.*

Some of the systems allow operation in a "simulation" mode whereby various structure or individual data type changes can be made and the results simulated. This is a distinct aid to the systems planner.

The most general and useful structure for this capability would be when a system would allow definitions in three categories, "Superseded," "Effective," and "Proposed" with only one Effective set of definitions being allowed concurrent with an unlimited number of Superseded or Proposed definitions. Naturally, the system should block any attempts to make changes in the Effective or Superseded modes.

9. *What copying facilities are available?*

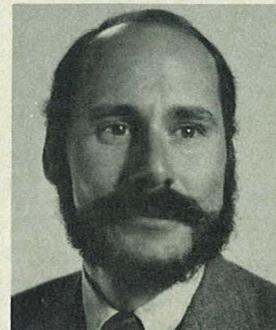
To be really appreciated by the programming staff, the ability to automatically generate data descriptions and COBOL, PL/1, or assembler code from existing definitions in the dictionary is important.

10. *What automatic controls are provided?*

Synonyms or several different names for the same definition may be all right and can be controlled by the dictionary. Homonyms or one name with different definitions are treacherous, however, and the package should enforce uniqueness here.

Conclusion

The 1970s have witnessed the emergence of data base and DBMS technology from the haven of the dp sophisticate into the real world. From now through the early 1980s, it is hard to believe that the use of data dictionaries will not follow a similar path as they become more appreciated in their role as controllers and as basic tools for analysis and design. Not only does it seem likely that most DBMS users will acquire data dictionaries for assistance in managing the corporate data resource, but many of the non-DBMS users will also analyze the advantages of a data dictionary for their environment. *



Mr. Schussel is vice president of the American Mutual Insurance Companies in Wakefield, Mass., where he is helping to use data base and data dictionary concepts for enhancing older non-flexible systems. Schussel is active on the Governor's Advisory Committee on Computers (Massachusetts) and has been Course Director of AMR's Data Base Design Course since 1971.