COMPUTER SCIENCE & TECHNOLOGY:

TECHNICAL PROFILE OF SEVEN DATA ELEMENT DICTIONARY/DIRECTORY SYSTEMS

500-3
7

NBS Special Publication 500-3
U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards

NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards[1] was established by an act of Congress March 3, 1901. The Bureau's overall goal is to strengthen and advance the Nation's science and technology and facilitate their effective application for public benefit. To this end, the Bureau conducts research and provides: (1) a basis for the Nation's physical measurement system, (2) scientific and technological services for industry and government, (3) a technical basis for equity in trade, and (4) technical services to promote public safety. The Bureau consists of the Institute for Basic Standards, the Institute for Materials Research, the Institute for Applied Technology, the Institute for Computer Sciences and Technology, the Office for Information Programs, and the Office of Experimental Technology Incentives Program.

THE INSTITUTE FOR BASIC STANDARDS provides the central basis within the United States of a complete and consistent system of physical measurement; coordinates that system with measurement systems of other nations; and furnishes essential services leading to accurate and uniform physical measurements throughout the Nation's scientific community, industry, and commerce. The Institute consists of the Office of Measurement Services, and the following center and divisions:

Applied Mathematics — Electricity — Mechanics — Heat — Optical Physics — Center for Radiation Research — Laboratory Astrophysics[2] — Cryogenics[2] — Electromagnetics[2] — Time and Frequency[2].

THE INSTITUTE FOR MATERIALS RESEARCH conducts materials research leading to improved methods of measurement, standards, and data on the properties of well-characterized materials needed by industry, commerce, educational institutions, and Government; provides advisory and research services to other Government agencies; and develops, produces, and distributes standard reference materials. The Institute consists of the Office of Standard Reference Materials, the Office of Air and Water Measurement, and the following divisions:

Analytical Chemistry — Polymers — Metallurgy — Inorganic Materials — Reactor Radiation — Physical Chemistry.

THE INSTITUTE FOR APPLIED TECHNOLOGY provides technical services developing and promoting the use of available technology; cooperates with public and private organizations in developing technological standards, codes, and test methods; and provides technical advice services, and information to Government agencies and the public. The Institute consists of the following divisions and centers:

Standards Application and Analysis — Electronic Technology — Center for Consumer Product Technology: Product Systems Analysis; Product Engineering — Center for Building Technology: Structures, Materials, and Safety; Building Environment; Technical Evaluation and Application — Center for Fire Research: Fire Science; Fire Safety Engineering.

THE INSTITUTE FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides technical services designed to aid Government agencies in improving cost effectiveness in the conduct of their programs through the selection, acquisition, and effective utilization of automatic data processing equipment; and serves as the principal focus wthin the executive branch for the development of Federal standards for automatic data processing equipment, techniques, and computer languages. The Institute consist of the following divisions:

Computer Services — Systems and Software — Computer Systems Engineering — Information Technology.

THE OFFICE OF EXPERIMENTAL TECHNOLOGY INCENTIVES PROGRAM seeks to affect public policy and process to facilitate technological change in the private sector by examining and experimenting with Government policies and practices in order to identify and remove Government-related barriers and to correct inherent market imperfections that impede the innovation process.

THE OFFICE FOR INFORMATION PROGRAMS promotes optimum dissemination and accessibility of scientific information generated within NBS; promotes the development of the National Standard Reference Data System and a system of information analysis centers dealing with the broader aspects of the National Measurement System; provides appropriate services to ensure that the NBS staff has optimum accessibility to the scientific information of the world. The Office consists of the following organizational units:

Office of Standard Reference Data — Office of Information Activities — Office of Technical Publications — Library — Office of International Standards — Office of International Relations.

---

[1] Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D.C. 20234.
[2] Located at Boulder, Colorado 80302.

DEC 28 1978
not Acc-Sing
QC 100
U57
no. 500-3
1977
c. 2

COMPUTER SCIENCE & TECHNOLOGY:

Technical Profile of Seven Data Element Dictionary/Directory Systems

Belkis Leong-Hong and
Beatrice Marron

Institute for Computer Sciences and Technology
National Bureau of Standards
Washington, D.C. 20234

U.S. DEPARTMENT OF COMMERCE,

Dr. Betsy Ancker-Johnson, Assistant Secretary for Science and Technology

NATIONAL BUREAU OF STANDARDS, Ernest Ambler, Acting Director

Issued February 1977

Reports on Computer Science and Technology

The National Bureau of Standards has a special responsibility within the Federal Government for computer science and technology activities. The programs of the NBS Institute for Computer Sciences and Technology are designed to provide ADP standards, guidelines, and technical advisory services to improve the effectiveness of computer utilization in the Federal sector, and to perform appropriate research and development efforts as foundation for such activities and programs. This publication series will report these NBS efforts to the Federal computer community as well as to interested specialists in the academic and private sectors. Those wishing to receive notices of publications in this series should complete and return the form at the end of this publication.

National Bureau of Standards Special Publication 500-3
Nat. Bur. Stand. (U.S.), Spec. Publ. 500-3, 45 pages (Feb. 1977)
CODEN: XNBSAV

Library of Congress Cataloging in Publication Data

Leong-Hong, Belkis.
Technical profile of seven data element dictionary/directory systems.

(Computer science & technology) (NBS special publication ; 500-3)
Supt. of Docs. no. : C 13.10:500-3
Bibliography: p.
1. Data base management. 2. Electron digital computers--Programming. I. Marron, B. A., joint author. II. Title. III. Series. IV. Series: United States. National Bureau of Standards. Special publication ; 500-3.
QC100.U57 no. 500-3 [QA76.9.D3] 602'.1s [001.6'442] 76-58915

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1977

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402
(Order by SD Catalog No. C13.10:500-3). Stock No. 003-003-01725-6 Price $1.05
(Add 25 percent additional for other than U.S. mailing).

PREFACE

Recognition of the need for automated tools to manage data has led to a variety of software products. This study was undertaken to provide a state-of-the-art assessment of one class of software tools, data element dictionary/directory systems.

This report identifies data element dictionary/directory systems and other computer facilities by trade names as necessary to provide a descriptive characterization of their features. This in no case implies a recommendation or endorsement by the National Bureau of Standards, nor should the presentation be construed as a certification that any system provides the indicated capabilities. The information presented was primarily obtained from vendors' documents, and has been reviewed by each vendor for accuracy and clarity, with the authors retaining the final technical judgement on the information included. The report is only intended to be informative and instructive, and not to be a competitive evaluation of systems for any specific application. Due to the dynamic nature of the systems' features, the information presented is current only to October 1976.

We gratefully acknowledge the excellent cooperation of the systems vendors, both in providing information and validating the technical content of this report. We especially appreciate the support of Dr. Dennis Fife, Chief, Computer Science Section, NBS.

CONTENTS

Page

1. INTRODUCTION 1

2. AUTOMATED TOOLS FOR THE MANAGEMENT OF DATA ELEMENTS 3

2.1 Background 3

2.2 Description of Automated Tools 6

3. DATA ELEMENT DICTIONARY/DIRECTORY 6

3.1 Characteristics of DED/D's 7

3.2 Classification of DED/D's 9

3.3 Primary DED/D 9

3.3.1 Free-Standing DED/D's 10
3.3.2 Dependent DED/D's 10
3.3.3 Primary DED/D Implementations 10
3.3.4 DED/D Relationship to DBMS 12

3.4 Secondary DED/D 13

4. SURVEY OF COMMERCIAL DED/D's 14

4.1 Survey Methodology 14

4.2 Scope-Notes for the Feature List 15

5. SUMMARY 16

CONTENTS (continued)

TABLES

Page

TABLE 1 Classification by Function 9

TABLE 2 List of Commercial DED/D's 11

TABLE 3 Relationship of DED/D to DBMS 12

APPENDICES

App. A Feature Analysis of Seven DED/D Systems 19

General Information 19
System Features-- Hardware Environment 20
System Features-- Software Environment 21
System Features-- User Environment 22
Data Element Attributes 23
Report Capabilities 25
Additional Output Capabilities 26

App. B Scope-Notes for the Feature List 27

App. C Glossary of Terms 31

App. D Bibliography 33

TECHNICAL PROFILE OF SEVEN DATA ELEMENT DICTIONARY/DIRECTORY SYSTEMS

Belkis Leong-Hong
and
Beatrice Marron

A Data Element Dictionary/Directory (DED/D) is a software tool that is used to control and manage data elements in a uniform manner. It can serve data base administrators, systems analysts, software designers, and programmers by providing a central repository for information about data resources across organization and application lines. This report describes and classifies DED/D systems and discusses the potential benefits from their use. A technical profile of seven commercially-available DED/D systems is presented with side-by-side exposition of technical features.

Keywords: Computer software; data base management system; data element dictionary; data element dictionary/directory; software tool.

1. INTRODUCTION

Data is a very valuable and sophisticated resource to an organization, not unlike the more traditional economic resources. Data is used to influence management decisions, by providing the decision-makers with timely and accurate information. Therefore, it is very important that data as a resource be easily accessible-- and it is essential that it be properly and effectively managed.

The explosive growth of databases, both in size and complexity, has made imperative the need for tools to aid in centrally controlling the database definitions and accesses; for tools to manage the growth and changes occurring in a database; and for tools to provide information to the different types of users within an organization. Data Base Management Systems (DBMS) have been designed and used to meet the information requirements for management of the organization.

A recent trend, however, is to use a separate class of automated tools for controlling/managing data elements in a uniform manner, across organizational lines. These automated tools, Data Element Dictionary/Directory (DED/D) systems, while performing some of the same functions as the DBMS, are different in that their main thrust is in providing control over ALL the data resources, automated and non-automated, within an organization.

In this report, seven commercially-available DED/D's are identified, and their features detailed. The systems included satisfy the following criteria:

1. The main function of the software package is to control and maintain the definitions and descriptions of the data elements.

2. The functions of defining, describing and controlling the data elements must be an integral part of the DED/D, i.e., it is not just a cross-referencing or report generating tool.

3. The software packages are commercially-available.

4. Documentation for the packages is available.

Time constraints did not permit a complete state-of-the-art survey. The systems presented are among those better known, and longer in operation. Inclusion of a system does not imply recommendation or endorsement. Similarly, omission of a system does not imply that its capabilities are less than those of an included system.

The seven systems chosen are:

DATA CATALOGUE (Synergetics Corporation)

DATA DICTIONARY (CINCOM Systems, Inc.)

DATA DICTIONARY/DIRECTORY with CMIS DATA ELEMENT GLOSSARY (IBM)

DATAMANAGER (MSP Inc.)

DB/DC DATA DICTIONARY SYSTEM (IBM)

LEXICON (Arthur Andersen & Co.)

UCC-TEN DATA DICTIONARY/MANAGER (University Computing Co.)

It is hoped that this report will assist data-conscious technologists by:

* Making them aware of the potential benefits derivable from the use of DED/D's;

* Providing at-a-glance technical characteristics of DED/D's;

* And by presenting descriptions of seven representative commercial systems.

2. AUTOMATED TOOLS FOR THE MANAGEMENT OF DATA ELEMENTS

In this section, a classification is given for the several types of automated tools available for the management of data elements.

Considerable ambiguity exists in the literature with respect to such terms as data dictionary, data directory, data administrator, data manager, etc. One source of confusion is in the use of the same terms to refer to both the automated function -- or software tool-- and to the human role associated with such functions, as for example, the data administrator or the data manager. The situation is further confused by the use of these same terms as names of proprietary packages, for example, DATAMANAGER and DATA DICTIONARY/DIRECTORY. (Software package names will be indicated by all upper-case letters, to distinguish them from descriptive terms).

To reduce this ambiguity, a glossary of terms used throughout this study is included as Appendix C. Whenever possible, terms commonly used in the area of data management are standard terms, as defined in The American National Dictionary for Information Processing, or in other standards documents. It may be premature at this point to standardize the usage of all terms in the area of the management of data elements because the field is changing so rapidly. A bibliography of relevant literature is included as Appendix D.

2.1 Background

Early design of data processing systems revolved around specific application systems; likewise, the data was organized so that it would be machine- and application-specific [23]. Thus, the data seldom crossed operational,

functional, or organizational boundaries. This situation resulted in multiple definition of the same data, as "independent" data files were generated-- thereby creating much redundancy, and overlap.

As the role of the computer grew within an organization, the need for system integration was evident, particularly with respect to data. The advent of Data Base Management Systems (DBMS) helped solve many of the information problems, by organizing the data elements to which they were applied. DBMS can be characterized as generalized software which provides a single flexible facility for accommodating different data files and operations, while demanding less programming effort than conventional programming languages [23]. Although the DBMS did not fully integrate all data resources within an organization, they helped unify many of those resources.

The use of DBMS can provide significant advantages [17]:

* The amount of redundancy in stored data can be reduced

* Problems of inconsistency in stored data can be avoided;

* Stored data can be shared;

* Standards can be enforced;

* Security restrictions can be applied;

* Data integrity can be maintained;

* Conflicting requirements can be balanced.

Over the years, data resources have grown in size and complexity. Then it was realized that not all of the data problems within the organization were necessarily resolved with the use of a DBMS, particularly since not all the data is automated. The proliferation of DBMS's, with varying degrees of sophistication and specialization [23], have resulted in situations where more than one DBMS is in use within a single establishment. These situations have dramatized the need for a facility that provides uniform and central control of all the data resources. The concept of centralization of control is growing in acceptance; usually the function of centralized control is performed by the Data Base Administrator (DBA) . The DED/D is a tool that helps the DBA in implementing this concept.

Many of the benefits realized from the use of a DED/D are parallel to the ones attributed to the use of a DBMS. However, it should be noted that while the benefits realized from a DBMS are directly related to the effective computer processing of the data; the benefits from a DED/D are directly related to the effective collection, specification and management of the total data resources of an organization.

Among the tangible benefits that can be derived from the use of a DED/D are:

* Simple and effective control of the data elements;

* Reduction of data redundancy and inconsistency;

* Enforcement of standard usage;

* Enforcement of security safeguards and controlled accessibility to the database;

* Determination of the impact on the total information activity from changes to data elements;

* Centralization of data elements as an aid in design and development of new systems.

* Consistency in documentation for data elements.

In a recently conducted survey, it was determined that 80% of those responding felt that the control over the data elements was a major advantage to the organization [33]. A U. S. General Accounting Office Report to the Congress [56] emphasized the need for government efforts to standardize data elements for computer systems. Such standardization could help reduce high cost of computer operations by eliminating unnecessary duplication and incompatibilities in collecting, processing and disseminating data.

A Federal Information Processing Standards Task Group has been organized to study Data Element Dictionaries. That group is surveying the DED/D's that are operational within the government, in a study which is parallel in scope to the subject of this report.

2.2 Description of Automated Tools

Automated tools for the management of data elements provide a centralized repository of information ABOUT each data element in order to facilitate access and control of the database. These tools do not manage the actual content of the data, but they do manage the descriptive characteristics of that data-- i.e., its physical properties, such as length, value range, types of admissible characters, and validation criteria. They also control the usage of the data elements-- i.e., what person or program is allowed to access and/or change the data elements. They further define the relationship of the data elements with each other, and with other components of the system. The use of automated software tools should reduce data redundancy, assure standard usage of data elements, and maintain data integrity.

These software tools are variously called catalogs, dictionaries, directories, and dictionary/directories. The deciding factor about what to call the package should be the amount and type of information that is provided the user. The following nomenclature is suggested:

*A data "catalog" simply lists all of the elements.

* A data element "dictionary" describes each data element, i.e., tells WHAT it is.

* A data element "directory" locates each data element, i.e., tells WHERE it is.

* A data element "dictionary/directory" describes and locates as well as lists each data element, i.e., tells both WHERE and WHAT it is.

Since most of the commercially-available packages are of the Data Element Dictionary/Directory (DED/D) type, focus will be placed on this group.

3. DATA ELEMENT DICTIONARY/DIRECTORY

The Data Element Dictionary/Directory is a software tool that provides the means for defining and describing the characteristics of a database, as opposed to the contents of a database. The basic features of a typical DED/D are described below. Then, a classification is imposed on the DED/D's, based on the kinds of functions they perform.

DED/D's are categorized by whether the dictionary/directory function is the "Primary" one in their use; or the function is "Secondary" in some other system (e.g. a DBMS). Primary DED/D's are further classified according to whether they are "Free-standing" or "Dependent" by implementation. Primary DED/D's are detailed, and several implementations are discussed.

3.1 Characteristics of DED/D's

Typically, a DED/D has a set of basic characteristics:

1) It contains a unique identification, a set of physical characteristics, and a textual description for each of the data elements;

2) It shows the relationships of elements to each other; and to components of the system, e.g., programs, reports.

3) It specifies the source, location, usage and destination of the elements.

4) It has validation and redundancy-checking capabilities;

5) It contains security safeguards to control the accessibility to the data elements;

6) It has a command language;

7) It has reporting capabilities, such as:

   a) Predefined management-oriented, statistical or summary reports;

   b) Ad-hoc user-defined reports;

   c) Cross-reference reports;

   d) Elements usage reports;

   e) Audit trail reports;

   f) Change-effect reports;

   g) Error reports;

8) It has retrieval capabilities, such as

keywording, indexing, and online or batch querying;

9) It has facilities for interacting with a DBMS.

The commercially-available packages have most of the above characteristics, as well as other features that distinguish each system from the rest. For example, some systems have online interactive capabilities, and some are predominantly batch systems. There is considerable difference in retrieval/reporting capabilities. A popular feature that is implemented in several DED/D's is the capability for generating data divisions or data-blocks for COBOL or PL/1 programs. Other distinguishing features include the automatic generation of program documentation, and the generation of program test data. Although most of the DED/D systems surveyed are implemented for large machines, there is at least one commercially-available DED/D that has been implemented also for a minicomputer.

The basic intent of DED/D's security features is to control the access to the data elements, however, this feature can also be used to protect the integrity of the database, as well as to enforce predetermined conventions. Security features are present in all DED/D's, although their extent and implementation vary from DED/D systems that provide their own security safeguards, including safeguards at the data element level, to those DED/D systems that make use of host system or operating system security provisions. In most cases, there is multi-level access control: the data administrator has the highest level of control--i.e., he has the ultimate authority for creating, updating, deleting and accessing all data elements-- and different levels of security are assigned to owners, and various classes of users of the data elements.

3.2 Classification of DED/D's

DED/D's can be grouped according to whether their dictionary/directory function is the primary one, or is secondary. Further, primary DED/D's can be subdivided into free-standing and dependent, according to their implementation. Table 1 presents a schematic representation of this classification.

Software Tool	Function	Implementation
DED/D	Primary	Free-Standing
		Dependent
	Secondary	Dependent (By definition)

Table 1
Classification by Function

3.3 Primary DED/D

A primary DED/D is a separate and distinct software package that functions MAINLY as a tool for identifying, locating, controlling, reporting and manipulating the information about data elements in a database. It is a basic tool within the database environment that can assist the data administrator, the systems designer/analyst, and the programmer in managing, planning, and evaluating the collection, storage, and usage of the data resources.

The existence of primary DED/D's as separate entities, rather than as a part of another system is a recent innovation in the area of data management. They may be implemented in such a way that they may require a DBMS to function properly. A further subdivision of the primary DED/D's elucidates the implementation of these packages: free-standing or dependent. It should be emphasized that both subcategories function principally as data dictionary/directory systems, and they are different only in their

implementation.

3.3.1 Free-Standing DED/D's. The free-standing DED/D's are self-contained, and perform the basic functions of controlling and managing the data elements without dependence on a DBMS. However, they may use programs not specifically written for the DED/D, in order to enhance their capabilities and performance.

A free-standing DED/D may support one or more DBMS through the use of interfaces, achieving mutual benefit from this association. It should be emphasized that the free-standing DED/D does not depend on the DBMS to function; however, the use of DBMS interfaces can provide the data base administrator with a greater degree of control over the DBMS. It is possible for free -standing DED/D's to have interfaces to more than one DBMS, sometimes simultaneously. Free-standing DED/D's are also known as "Generalized DED/D's". [33]

3.3.2 Dependent DED/D's. The dependent DED/D's are separate software systems that are specifically tailored to a general purpose DBMS, and provide the DBMS with control and management of the data elements by supplying the DBMS with the description, definition, location, and cross-references of the constituent data elements. In turn, DBMS resources such as file structure and access methods are made available to the DED/D. Dependent DED/D's must perform all the functions of free-standing DED/D's.

Because the dependent DED/D is designed and implemented to be DBMS-specific, the portability of this type of DED/D is restricted to installations having that particular DBMS.

3.3.3 Primary DED/D Implementations. Of the seven commercially-available primary DED/D systems included in this report, four are free-standing, and three are dependent. Table 2 provides the names of the software packages and the vendor's name; the implementation classification of the DED/D; the hardware on which it is implemented; and the operational mode of the system.

COMMERCIAL DED/D	TYPE	HARDWARE IMPLEMENTATION	OPERATIONAL MODE
DATA CATALOGUE (Synergetics Corporation)	Primary Free-standing	I360/370 U1108	Batch Online
DATA DICTIONARY (CINCOM SYSTEMS, Inc.)	Primary Dependent	I360/370	Batch
DATA DICTIONARY/DIRECTORY with CMIS DATA ELEMENT GLOSSARY (IBM)	Primary Free-standing	I360/370	Batch
DATAMANAGER (MSP Inc.)	Primary Free-standing	I360/370	Batch Online with TSO, CICS and CMS
DB/DC DATA DICTIONARY SYSTEM (IBM)	Primary Dependent	I360/370	Batch Online with IMS/DC
LEXICON (Arthur Andersen & Co.)	Primary Free-standing	I360/370 IBM System 3	Batch Online with TSO or IMS/DC
UCC-TEN DATA DICTIONARY/MANAGER (University Computing Co.)	Primary Dependent	I360/370	Batch Online with IMS/DC

Table 2
List of Commercial DED/D's

3.3.4 DED/D Relationship to DBMS. Because DED/D's are concerned with the management of data elements, it is logical that there should exist a strong relationship between a DED/D and a DBMS. In fact, in this report, DED/D's are categorized as "Free-standing", with the capability to interface with a DBMS; or as "Dependent" on a DBMS.

Table 3 shows the relationship between DED/D's and DBMS's. On the vertical axis are the seven commercially-available DED/D's included in this study, and on the horizontal axis are the DBMS that the DED/D's are related to. The entries consist of either an "I" to indicate that an interface exists, or a "D" to indicate that the DED/D is dependent on the given DBMS.

	Data Base Management Systems				
DED/D Systems	ADABAS	IMS	MARK IV	TOTAL	IDMS
DATA CATALOGUE		I	I	I	
DD/D (IBM)					
DD (CINCOM)				D	
DATAMANAGER	I	I	I	I	I
DB/DC DDS		D			
LEXICON		I		I	I
UCC-TEN		D			

Table 3
Relationship of DED/D to DBMS
(I = Interface exists; D = Dependency)

It should be noted that free-standing DED/D's can have interfaces to more than one DBMS, as shown in Table 3. Moreover, these interfaces can be supported simultaneously, so that where more than one DBMS is operational, the DED/D having multiple interfaces can generate the data definitions for all of the DBMS's, without having to "disconnect" any of the interfaces. The implementation of interfaces varies considerably from system to system. The variance may be due to the way that the control blocks are generated, or it may be

in the way that the DED/D supports the DBMS.

Interfaces enhance the usefulness of both the DBMS and the DED/D, by providing the user with the ability to:

1) Define the database to the DED/D, capitalizing on DBMS resources, such as existent, well-defined file structures and access methods;

2) Generate data element definitions for a DBMS from an up-to-date DED/D;

3) Exercise control over the data elements of a DBMS using DED/D facilities.

3.4 Secondary DED/D

A secondary DED/D is a software package in which the data dictionary function exists, but is not the "raison d'etre" of the software. Secondary DED/D's usually are embedded functions within another system; they serve as the data and file pre-definition mechanism, and are an integral physical part of another system.

Among the major differences between a primary and a secondary DED/D are:

1) The primary DED/D is a self-contained system, whereas the secondary DED/D is an internal function of another software system;

2) The reporting and retrieval capabilities are extensive for the primary DED/D, and modest for the secondary;

3) Primary DED/D's have more extensive security control over the data elements.

Since secondary DED/D's are embedded within another system, they are necessarily oriented towards the characteristics and internal representation of that system. It is impossible to separate their functions from those of the software system they serve. The emphasis of this report is on primary DED/D's; secondary DED/D's are included only to complete the classification.

4. SURVEY OF COMMERCIAL DED/D's

Several commercially-available primary DED/D's were identified and analyzed. Details of this study are presented in this section and in Appendix A, in order to aid potential users in the appraisal of these software packages.

4.1 Survey Methodology

Seven systems were selected based on a survey of the literature [especially 1, 8, 16, 32, 33], and on the criteria outlined earlier. The hardware on which the DED/D systems were implemented was not a factor in the choice. It must be reiterated that no recommendation or endorsement of these systems is implied. This field is so new and dynamic that new systems, and new versions of systems, are appearing with increasing frequency. In fact, several other eligible systems have come to our attention too late to be included; one selected system had to be excluded when it was withdrawn from the marketplace.

Once the applicable systems were chosen, the vendors of these packages were contacted, and detailed technical documentation was obtained for each system. To insure uniformity of coverage, it was deemed necessary to establish a consistent set of descriptors for system features. These descriptors constitute the Feature List, and they are the basis for the technical profile of the seven systems presented side by side in Appendix A. A set of scope-notes was developed for these descriptors, and it is included as Appendix B.

The information used in the system descriptions was obtained from vendor documentation. In cases where the available documentation did not provide sufficient data, further information was obtained by direct consultation with the developers. Every effort has been made to ensure the accuracy and timeliness of the data; the vendors were asked to validate the data, and their concurrence was obtained for the system descriptions. However, omissions and out-of-date information are inevitable in a survey of such an extremely active and evolutionary field as this. In all cases, only information about operational rather than planned capabilities is included.

4.2 Scope-Notes for the Feature List

The Feature List is organized under the following seven major headings:

1) General Information

   Information of a general nature is included in this section.

2) System Features-- Hardware Environment

   Information on the hardware environment in which the DED/D is implemented.

3) System Features-- Software Environment

   Information about the software requirements for the DED/D.

4) System Features-- User Environment

   Information about system facilities available to the user for definition, access, and retrieval of the content of the DED/D.

5) Data Element Attributes

   Information about the characteristics of a data element as represented in a DED/D.

6) Report Capabilities

   Typical reports that can be produced by the DED/D are listed in this section.

7) Additional Output Capabilities

   Other DED/D outputs and reports are listed in this section.

It should be noted that although DED/D's can provide control for several levels of data (data elements, groups, files, programs, systems), this analysis concentrates on the data element level. Each DED/D treats the various data levels differently, but there is commonality in the treatment at the most basic level, the data element level.

The scope-notes for the Feature List are reproduced in Appendix B, following the side-by-side descriptions of the seven systems. It should be emphasized that the Feature List represents an objective reporting of technical

characteristics of each system, and it is not meant to convey critical or subjective system comparisons.

5. SUMMARY

This report identifies technical features of Data Element Dictionary/Directory (DED/D) systems, and presents a side-by-side description of seven representative commercially-available systems. It is hoped that this feature-list approach will aid prospective users by providing a mechanism for matching data needs against DED/D system capabilities.

Among the major advantages of utilizing a DED/D are:

* A central, consistent source of information about data is provided to an organization.

* Consistent and timely documentation about data resources is provided to the manager.

* A tool for controlling and maintaining the data resources of an organization is provided to the data base administrator.

* A tool for database and application design is provided to the system analyst and designer.

Some disadvantages of utilizing a DED/D are: it can be time-consuming to install; the maintenance function may require considerable effort; and there may be objection to the formality necessitated by the DED/D.

After weighing these advantages and disadvantages in the light of his organization's information requirements, a manager can make the decision whether or not to obtain a DED/D. If the decision is in favor of a DED/D, then he has basically two options-- he can design and implement an in-house system, or he can purchase/lease a commercial system.

The first choice implies a high initial cost in both time and labor. Another drawback is that such a system is usually tailored to a specific application, and may not be flexible and responsive to changing requirements. However, in spite of these problems, many data managers realizing the advantages of using a DED/D, have opted to develop their own system when commercially-developed ones were not available [33].

The purchase/lease option generally offers a shorter period of time required for development of a working system. Less manpower is usually required to install and support an off-the-shelf package, than to develop the software from scratch, thereby enabling better utilization of personnel resources for problem-oriented analysis rather than detailed coding of the DED/D system. In addition, commercial systems are generally more versatile across applications.

Sometimes the choice of a commercial system may be influenced by the fact that a DBMS is already operational within the organization. If there exist DED/D's tailored to that DBMS, the manager can consider dependent or free-standing DED/D's having interface capability to that DBMS. Choice may also be influenced by the degree of support given a package. Of course, cost, availability, and specificity are critical factors in the purchase/lease of a DED/D system.

Several observations can be made from this study:

* Key features of DED/D's include control at the data element level, element usage standardization, extensive reporting capabilities, and interaction with application programs.

* There is a strong relationship between the use of a DED/D and the use of a DBMS.

* The current implementations are largely limited to one major equipment configuration.

* The purchase price of the systems studied ranges from $10-15K. (This is approximately 1/4 the purchase price of a DBMS.)

* Vendor support varies from practically no support, to full support, including the provision of a forum for user's information exchange.

* The most common application for DED/D's has been in industries that have a propensity for very large databases, such as manufacturing, utilities, banking, and insurance [43,44]. However, as databases have become larger and more complex in other industries, DED/D's are being more widely used as a basic data management tool [see 2, 14, 20, 29, 40, 48, 50, 59].

* The availability of DED/D's has facilitated the standardization of data element definitions for industry-wide use. One of the commercial DED/D's provides a glossary of common manufacturing terms,

which helps in establishing uniformity in industry-wide usage of data elements. An example of profession-wide data element standardization is in clinical medicine, where a Medical Data Elements Lexicon [24] has been suggested for clinical medical records.

* DED/D's are of considerable use in auditing, because they facilitate transaction tracing, and provide security information at the data element level. They also aid in the development of compliance testing and audit software [1, 9].

* The market for DED/D's has lagged the market for DBMS, but there is growing awareness of DED/D's.

No comprehensive statistics are available on the use of DED/D's, but one survey projects that they will be "a standard feature in almost all very large installations by 1978" [33]. An area for future study is users' experiences with the various systems. Another aspect for investigation is the kind of usage of DED/D's. Information about whether such systems are being used primarily as reporting/documentation tools, as preprocessors for DBMS, or for change-effect projections in decision-making, can be valuable to the information systems community.

SOFTWARE PACKACE NAME	DATA CATALOCUE	DATA DICTIONARY	GENERAL INFORMATION DATA DICTIONARY/ DIRECTORY WITH CMIS DATA ELEMENT CLOSSARY	DATAMANAGER	DB/DC DATA DICTIONARY SYSTEMS	LEXICON	UCC TEN DATA DICTIONARY & MANACER
VENDOR	Synergetics Corporation	Cincom Systems, Inc.	IBM	MSP Inc.	IBM	Arthur Andersen & Co.	University Computing Company
ADDRESS	1 De Angelo Dr. Bedford, MA 01730	307 Maples Ave, W. Vienna, VA 22180	Local IBM Branch Office	133 Mt. Auburn St. Cambridge, MA 02138	Local IBM Branch Office	69 W. Washington St. Chicago, IL 60602	Commercial Div. P. O. Box 47911 Dallas, TX 75247
CONTACT NAME AND TELEPHONE NUMBER	Alexander De Vito 617-275-0250	Norman L. Bushee 703-281-2121	Local IBM Branch Office Marketing Representative	Vice-President of Sales 617-492-3316	Local IBM Branch Office Marketing Representative	Richard Nerad 312-346-6262	Local UCC TEN Representative, or Products Manager (Dallas) 214-688-7174
PURCHASE COST	$12.9K (DOS) $14.9K (OS)	$11.5K original $9.5K subsequent copies	Not Applicable	Modular prices from $9.9K for perpetual license	Not Applicable	Not Applicable	$16.5K (CSA Schedule for Federal Agencies)
LEASE COST	Negotiable	$275/mo.	Perpetual lease: $523/mo. for 1 yr., no charge after 1st year	Monthly rental: 6.25% of purchase price (minimum 1 year): Full pay-out lease available	Program Product OS/VS: $580/mo. DOS: $290/mo.	$9K for Data Dictionary; $9K for Input Processor and Report Writer	$650/mo. for 3 yrs Lease-dependent terms available
MAINTENANCE COST	$1K (DOS); $1.2K (OS) after 1st year	$500/year (single use charge)	No maintenance support (Installed User Program)	10% of current sales price after 1 year; no charge if leased	Part of the license fee (Class B maintenance)	Maintenance fee included in lease	$2K after 1 year (if purchased); lease cost includes maintenance
PROPRIETARY ?	Yes	Yes	Yes	Yes	Yes	Yes	Yes

SYSTEM FEATURES - HARDWARE ENVIRONMENT

SOFTWARE PACKAGE NAME	DATA CATALOGUE	DATA DICTIONARY	DATA DICTIONARY' DIRECTORY WITH CMIS DATA ELEMENT GLOSSARY	DATAMANAGER	DB/DC DATA DICTIONARY SYSTEMS	LEXICON	UCC TEN DATA DICTIONARY & MANAGER
HARDWARE IMPLEMENTATION	IBM 360/370 (Also Univac 1100)	IBM 360/370	IBM 360/370	IBM 360/370	IBM 370	IBM 360/370 (Also IBM System 3)	IBM 360/370
OPERATING SYSTEM	OS; OS/VS; DOS; DOS/VS EXEC-8	All IBM 360/370	OS/MFT; OS/MVT	OS; OS/VS; DOS; DOS/VS	OS/VS; DOS/VS	OS/MFT; OS/MVS; VS; DOS/MVT	OS; OS/VS
CORE REQUIREMENT	100-125K (IBM) 25-35K (UNIVAC)	64-100K*	128K for building and changing; 192K for printing	48K under OS; 52K under DOS	(virtual storage) IMS/VS:* online: 800K batch: 600K DL/1 DOS/VS:* 600K + VSAM req.	256K	Online: 70K* Batch: 120K*
PERIPHERALS	1 input device 2 disk drives 1 printer 1 terminal for online oper.	1 disk drive 1 card reader 1 printer	4 tape drives 2 disk drives 1 card reader/ punch 1 printer	1 tape drive 1 disk drive 1 printer 1 terminal for online oper.	IMS/VS hardware requirements, plus I-3270 Display terminal required for online operation	1 card reader/ punch 1 tape device 1 disk drive 1 printer 1 terminal for online oper.	1 disk drive 1 card reader 1 printer I-3270 for online operation
OPERATIONAL MODE	Batch and online	Batch	Batch	Batch, and online via TSO, CICS, and CMS	Batch, and online for IMS/VS	Batch, and online via TSO or IMS/DC	Batch and online
RUNNING ENVIRONMENT	Test and Production	Test and Production	Production	Test and Production	Test and Production	Test and Production	Test Production

*The core requirement specified is in addition to requirements for the host system.

SOFTWARE PACKAGE NAME	DATA CATALOGUE	SYSTEM FEATURES - SOFTWARE ENVIRONMENT DATA DICTIONARY	DATA DICTIONARY/ DIRECTORY WITH CMIS DATA ELEMENT GLOSSARY	DATAMANAGER	DB/DC DATA DICTIONARY SYSTEMS	LEXICON	UCC TEN DATA DICTIONARY & MANAGER
PROGRAMMING LANGUAGE	ANS COBOL	ALC	PL/1 level F	ALC	ALC	ALC and COBOL	ALC and COBOL
USE OF OTHER PROGRAMS	None	"User-Exit" facility and IBM sort	Interface program for user program data definition; Input stream handler; and PL/1 sort/merge	IBM System routines	System Control Programs(SCP); IMS/VS or DL/1 DOS/VS programs	IMS file definition utilities	IMS utilities
USE OF DBMS	Interfaces to IMS, IMS/DL1, MARK IV, and TOTAL	Requires TOTAL	None	Interfaces to ADABAS, IDMS, IMS, IMS/DL1, MARK IV, and TOTAL	Requires IMS/VS or DL1 DOS/VS	Interfaces to IDMS, IMS, and TOTAL	Requires IMS/360 or IMS/VS
FILE STRUCTURE	ISAM, VSAM	TOTAL file structures	ISAM	BDAM	IMS file structures	OS file structures; IDMS, IMS, and TOTAL file structures	OS and IMS file structures
FILE UPDATE	Single-entry	Single-entry for elements; individual for aliases	Single-entry for non-variable info.; individual for variable info.	Single-entry	Single-entry and Individual	Single-entry	Single-entry
FILE BACKUP AND RECOVERY	DATA CATALOGUE utilities	IBM and TOTAL utilities	IBM utilities	DATAMANAGER utilities	IMS utilities	IBM utilities	IMS utilities

SYSTEM FEATURES - USER ENVIRONMENT

SOFTWARE PACKAGE NAME	DATA CATALOGUE	DATA DICTIONARY	DATA DICTIONARY/ DIRECTORY WITH CMIS DATA ELEMENT GLOSSARY	DATAMANAGER	DB/DC DATA DICTIONARY SYSTEMS	LEXICON	UCC TEN DATA DICTIONARY & MANAGER
INPUT LANGUAGE	Fixed-form definition language; Free-form definition language (English-like); From application programs	Fixed-form "initial load"; Free-form keyword-oriented maintenance	Free-form and Fixed-form ("Data Specification Methodology"); From application programs	Free-form definition language	Free-form command language Fixed-form "Interactive Display Forms"; DBMS data definitions; From application programs	Free-form definition language (English-like)	Fixed-form ; keyword-oriented ; 3270 "menu"
COMMAND LANGUAGE	Fixed-form and Free-form	Free-form	Through "Data Specification Methodology"	Free-form	Free-form	Free-form Report Definition Language	Free-form transaction-driven
ONLINE QUERIES?	Yes	No	No	Yes	Yes	Yes	Yes
ONLINE QUERY LANGUAGE	Free-form; Query-through-indices	Not applicable	Not applicable	Free-form	Fixed-form thru "Interactive Display Forms" facility	Free-form	Free-form keyword; thru 3270- "Menu"
TUTORIAL?	Yes, help command	No	No	No	Yes, thru 3270 "Interactive Display Forms" facility	Yes	Yes, thru 3270-"Menu"
OUTPUT FORM	Hard copy and Microform	Hard copy	Hard copy	Hard copy and Microform	Hard copy	Hard copy and Microform	Hard copy

DATA ELEMENT ATTRIBUTES

SOFTWARE PACKAGE NAME	DATA CATALOGUE	DATA DICTIONARY	DATA DICTIONARY/ DIRECTORY WITH CMIS DATA ELEMENT GLOSSARY	DATAMANAGER	DB/DC DATA DICTIONARY SYSTEMS	LEXICON	UCC TEN DATA DICTIONARY & MANAGER
ELEMENT IDENTIFIER	Max 20 chars. for control field; 30 chars. COBOL name	1-30 A/N characters	6 chars. (left just., 1st alpha); or 30 chars. label-assigns 6 chars. pseudo-label	Max. 32 chars.	Max. 31 chars.; 3 qualifiers (status, occurence, code) required	COBOL, PL/1, or ALC compilable names	1-8 A/N characters followed by unique "Definition Number"
SYNONYM?	Yes, 4000 allowed	Yes, restricted to 1-30 A/N chars.	Yes	Yes, 16 allowed, max. 80 chars.	Yes	Yes	Yes, 255 allowed
KEYWORD CAPABILITY?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DATA STRUCTURE	Hierarchical	Network	Hierarchical	Hierarchical	Hierarchical	Hierarchical, Network, and Relational	Hierarchical
CHARACTER TYPE	Alpha, numeric, alphanumeric, and special	Alpha, numeric, alphanumeric, and special	Alpha, numeric, alphanumeric, and special	Alpha, numeric, alphanumeric, and special	Alpha, numeric, alphanumeric, and special	Alpha, numeric, alphanumeric, and special	Alpha, numeric, alphanumeric, and special
LENGTH	May be specified	May be specified; required for SOCRATES and DDL generator elements	No requirement	May be specified	May be specified	Max. 500 chars.	May be specified
JUSTIFICATION?	Yes, may be specified: J=justified S=synchronized	No	Yes, may be specified, right or left	No	Yes, may be specified	No	Yes, may be specified, right or left
VALUE RANGE?	Yes	No	Yes	Yes	Yes	No	No

DATA ELEMENT ATTRIBUTES - Continued

SOFTWARE PACKAGE NAME	DATA CATALOGUE	DATA DICTIONARY	DATA DICTIONARY/ DIRECTORY WITH CMIS DATA ELEMENT GLOSSARY	DATAMANAGER	DB/DC DATA DICTIONARY SYSTEMS	LEXICON	UCC TEN DATA DICTIONARY & MANAGER
VALIDATION	DED/D facility	DED/D facility	DED/D facility	DED/D facility	DED/D (limited) and User-written	DED/D facility	DED/D facility and user-written
REDUNDANCY/ INCONSISTENCY CHECK?	Yes, DED/D facility	Yes, DBMS facility	Yes, DED/D facility	Yes, DED/D facility	Yes, DED/D facility	Yes, DED/D facility	Yes, DED/D facility and "software switch"
DEFINITION/ DESCRIPTION?	Yes, narrative description (1-5000 chars) or keyword notation	Yes, narrative description	Yes, free-form narrative description	Yes, narrative description (max. 65,000 lines/elem.)	Yes, narrative description; also thru "User Data"	Yes, narrative description	Yes, narrative description; also summary definition (1-30 chars.)
RELATIONSHIPS?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
OWNER/USER?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SECURITY SAFEGUARDS	DED/D, optional feature: 3 types of passwords; security at element level may be specified	DED/D, password facility; security at element level may be specified; additional security available thru "user exit"	DED/D, security at element level may be specified	DED/D, protection, authority, and password facility; security at element level may be specified; some commands available only to DBA	Thru IMS and operating system: IMS/VS security subsystem; segment sensitivity (PSB); operating system security (VSAM)	DED/D special security module; security at element level may be specified; 3 levels available; also requires user-supplied password	Thru IMS and operating system

REPORT CAPABILITIES

SOFTWARE PACKAGE PACKAGE NAME	DATA CATALOGUE	DATA DICTIONARY	DATA DICTIONARY/ DIRECTORY WITH CMIS DATA ELEMENT GLOSSARY	DATAMANAGER	DB/DC DATA DICTIONARY SYSTEMS	LEXICON	UCC TEN DATA DICTIONARY & MANAGER
DATA DICTIONARY/ DIRECTORY	Yes	Yes	Yes	Yes	Yes	Yes	No
MANAGEMENT-ORIENTED	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AD-HOC	Yes	Yes	No	Yes	No	Yes	No
ELEMENT USAGE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CROSS REFERENCE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
KWIC or KWOC INDICES	Yes, both	No	Yes, both	Yes, both	No	Yes, both	Yes, both
AUDIT TRAIL	Yes	Yes	Yes	Yes	Yes (IMS audit trail is used)	Yes	Yes
CHANGE-EFFECT	Yes	Yes	No	Yes	Yes	Yes	Yes
ERROR	Yes	Yes	Yes	Yes	Yes	Yes	Yes

ADDITIONAL OUTPUT CAPABILITIES

SOFTWARE PACKAGE NAME	DATA CATALOGUE	DATA DICTIONARY	DATA DICTIONARY/ DIRECTORY WITH CMIS DATA ELEMENT GLOSSARY	DATAMANAGER	DB/DC DATA DICTIONARY SYSTEMS	LEXICON	UCC TEN DATA DICTIONARY & MANAGER
DATA AREAS FOR PROGRAMS?	Yes, for COBOL, PL/1, Assembly, MARK IV	No	Yes, for COBOL, PL/1	Yes, for COBOL, PL/1, Assembly	Yes, for COBOL, PL/1, Assembly	Yes, for COBOL, PL/1, ALC	Yes, for COBOL, PL/1, Assembly
PROGRAM DOCUMENTATION ?	No	No	No	Yes	Yes	Yes	Yes
OTHER OUTPUTS	Update transactions from user's programs; DBD & PSB for DL/1; PCB masks; SSA; DBGEN for TOTAL; generates control statements for MARK IV Query Retrieval report.	Generate Data Definition Language (DDL) for TOTAL; SOCRATES Label Dictionary conversion facility.	Manufacturing Glossary (CMIS Data Elements Glossary-DEG); Record layout.	Data descriptions for ADABAS, DL/1, IDMS, IMS, SYSTEM 2000, TOTAL; Display actual content of source files; System documentation; Test file generation.	DBD & PSB for IMS/VS & DOS/DLI Interactive Display Forms Facility; Inquiry Reports using GIS.	Control blocks for IDMS, IMS, TOTAL; DBD & PSB for DL/1; PCB masks; SSA; Program linkage section; DBMS-generation description report; DD & RD Dictionary maintenance Report; segment I/O area.	Program profiles; Record layout; Master Terminal Operator Report; 3270 message format control blocks; IMS STAGE 1 SYSGEN; Sysgen compare report; SSA; PCB masks; DBD & PSB for DL/1 Terminal security maintenance.

APPENDIX B

SCOPE-NOTES FOR
THE FEATURE LIST OF
DATA ELEMENT DICTIONARY/DIRECTORY (DED/D) SYSTEMS

(The Feature List is designed for "yes/no" and short answers)

GENERAL INFORMATION

SOFTWARE PACKAGE NAME: Name or acronym by which the DED/D is known.

VENDOR: Name of the company that markets the product.

ADDRESS: Address of the company that markets the product.

CONTACT NAME & TELEPHONE NUMBER: Name and phone number for further information.

PURCHASE COST: Cost for purchase of software package, or perpetual license to use the package.

LEASE COST: Cost for lease of software package, and period of the lease.

MAINTENANCE COST: Cost for service or maintenance contract, and period of contract.

PROPRIETARY ? Proprietary status of the package.

SYSTEM FEATURES-- HARDWARE ENVIRONMENT

HARDWARE IMPLEMENTATION: The primary hardware implementation of the DED/D. (Also include other operational implementations).

OPERATING SYSTEM: The operating system under which the DED/D operates.

CORE REQUIREMENT: The minimum amount of memory required to use the DED/D.

PERIPHERALS: Peripheral equipment required to use the DED/D.

OPERATIONAL MODE: Mode in which the DED/D operates, i.e., batch or online.

RUNNING ENVIRONMENT: The "running modes of the DED/D, i.e., test mode or production mode.

SYSTEM FEATURES-- SOFTWARE ENVIRONMENT

PROGRAMMING LANGUAGE: The programming language in which the DED/D software is written.

USE OF OTHER PROGRAMS: Note required programs not specifically written for the DED/D.

USE OF DATA BASE MANAGEMENT SYSTEM (DBMS): Note interfaces to, or requirements for, a DBMS.

FILE STRUCTURE: File structure supported by the DED/D.

FILE UPDATE: Kind of file updating capability for the DED/D, i.e., "Single-entry" update, (the DED/D automatically updates every occurrence of the affected entry), or "individual" update, (only the specified occurrence is updated, and other occurrences have to be updated individually).

FILE BACKUP AND RECOVERY: Source of DED/D system backup and recovery capabilities.

SYSTEM FEATURES-- USER ENVIRONMENT

INPUT LANGUAGE: Does the input language require fixed-form or free-form inputs? Note other methods of input.

COMMAND LANGUAGE: Does the command language require fixed-form or free-form commands? Note other methods.

ONLINE QUERIES? Does the DED/D have online query capability?

ONLINE QUERY LANGUAGE? Does the query language require fixed-form or free-form queries? Note other methods.

TUTORIAL? Is there capability for tutorials, "help," or "prompting" to guide the user?

OUTPUT FORM: Physical form of the DED/D system output, i.e., hard-copy (any printed output) or microform.

DATA ELEMENT ATTRIBUTES

ELEMENT IDENTIFIER: Provide restrictions placed on the name of the data element (identifier or designator used for referencing elements contained in the DED/D.)

SYNONYM ? Can a synonym (word or abbreviation) be used as a substitute for the data element identifier? Note any restrictions.

KEYWORD CAPABILITY ? Can an element be identified as a keyword?

DATA STRUCTURE: Data structure supported by the DED/D, i.e., hierarchical, network, relational.

CHARACTER TYPE: Types of characters which can be used in the representation of the content of data elements.

LENGTH: Can the size or length of the representation of the content of data elements be specified?

JUSTIFICATION? Can justification be specified for the representation of the content of data elements?

VALUE RANGE? Does the DED/D provide for a range of value in the representation of the content of data elements?

VALIDATION: Indicate if data validation (edit) is performed by the DED/D or by other means.

REDUNDANCY/INCONSISTENCY CHECK? Is there a specific feature that identifies whether the data element is redundant or inconsistent? Indicate whether it is a DED/D or other facility.

DEFINITION/DESCRIPTION ? Is there a facility for defining and/or describing a data element in narrative form?

RELATIONSHIPS ? Is there a capability for specifying the relationship of a data element to another data element or to a higher level structure?

OWNER/USER ? Is there a facility for indicating authorized users (persons or programs) or owners of data elements?

SECURITY SAFEGUARDS: Indicate if security safeguards are provided by the DED/D or by other means.

REPORT CAPABILITIES

DATA DICTIONARY/DIRECTORY: Is a listing of the entire DED/D one of the system's formal report?

MANAGEMENT-ORIENTED: Does it produce management-oriented, summary, and/or statistics reports?

AD-HOC: Can the user structure his own reports, on an ad-hoc basis?

ELEMENT USAGE: Does it produce reports that indicate where and how the data elements are used?

CROSS-REFERENCE: Does it produce cross-reference and/or relationship reports?

KWIC OR KWOC INDICES: Does it produce indices based on keyword-in-context (KWIC), or keyword-out-of-context (KWOC)?

AUDIT TRAIL: Does it produce audit trail reports?

CHANGE-EFFECT: Does it produce reports that tell what effect a change in data element(s) would have on the DED/D system?

ERROR: Does it produce error reports?

ADDITIONAL OUTPUT CAPABILITIES

DATA AREAS FOR PROGRAMS? Can the DED/D produce the data division or data blocks for user application programs? If yes, in what language?

PROGRAM DOCUMENTATION? Does the DED/D have the facility to produce documentation for user application programs, based on the data definition or description of the data elements?

OTHER OUTPUTS: Include here any other output/report capabilities of the DED/D.

APPENDIX C

GLOSSARY OF TERMS

Automated Tool: (See Software Tool)

Database (or Data Base): (1) A collection of data consisting of at least one file, that is sufficient for a given purpose or for a given data processing system; (2) A collection of data fundamental to a system or to an enterprise.

Data Administrator: (see Data Base Administrator)

Data Base Administrator (DBA): A person or group of people that has the functions of centrally managing, controlling, and organizing the database for an organization. The responsibilities of the DBA include the control, definition, organization, documentation, protection, and efficiency of the database.

Data Base Management System (DBMS): A Data Base Management System can be characterized as a generalized software tool that provides a single, flexible facility for accommodating different data files and operations, while demanding less programming effort than conventional programming languages. A DBMS facilitates operations on data (definition, maintenance, storage, retrieval, output); it facilitates reference by name rather than by physical location; and it provides an environment that is not tied to a particular set of application programs or files.

Data Catalog: A software tool used to list all of the data elements in a database.

Data Dictionary: (See Data Element Dictionary)

Data Dictionary/Directory: (See Data Element Dictionary/Directory)

Data Directory: (See Data Element Directory)

Data Element Dictionary: A software tool used to describe each data element, i.e., to tell "what" it is.

Data Element Dictionary/Directory (DED/D): A software tool used to list, describe, and locate each data element in a database. It provides a centralized repository of information about each data element in order to facilitate management and control of, and access to the

database.

Data Element Directory: A software tool used to locate each data element, i.e., to tell "where" it is.

Data Manager: (See Data Base Administrator)

Dependent DED/D: A primary DED/D that is designed and implemented to be DBMS-specific. It uses features of the DBMS it is tailored to, while providing the DBMS with control and management of the data elements.

Free-standing DED/D: A self-contained DED/D that performs the basic functions of controlling and managing the data elements without dependence on a DBMS.

Generalized DED/D: (See Free-standing DED/D)

Primary DED/D: A separate and distinct software package that functions principally as a tool for identifying, locating, controlling, reporting and manipulating the information about data elements in a database.

Secondary DED/D: A software package in which the data dictionary function exists, but is not the main purpose of the software. A secondary DED/D is usually embedded in another system, and serves as the file and predefinition mechanism for that system.

Software tool: A computer program, rules, and associated documentation that assists a data processing technologist in designing, developing, maintaining and managing data and software.

APPENDIX D

BIBLIOGRAPHY

[1] Adams, Donald L., editor, "System and Audit Aspects of the Data Dictionary," EDPACS, vol. III, no. 11, May 1976.

[2] Albert, T.M., "Control of Logistics Data in the Defense Department," First National Symposium on the Management of Data Elements in Information Processing, January 1974, pp. 1-12.

[3] Army Material Command, "AMC Data Element Dictionary," AMCP 18-1, vol.2, January 1971, 7 pages.

[4] Batman, R.B., "End User Data Control," First National Symposium on the Management of Data Elements in Information Processing, January 1974, pp. 13-28.

[5] Bontempo, Charles J. and David G. Swanz, " Data Resource Management," First National Symposium on the Management of Data Elements in Information Processing, January 1974, pp. 29-36.

[6] Cahill, John J., "A Dictionary/ Directory Method for Building a Common MIS Data Base," Journal of Systems Management, November 1970, pp. 23-29.

[7] Canning, Richard G., editor, "The Current Status of Data Management," EDP Analyzer, vol. 12, no. 2, February 1974.

[8] Canning, Richard G., editor, "The Data Dictionary/Directory Function," EDP Analyzer, vol. 12, no. 11, November 1974.

[9] Chick, Morey J., "The Cost of Information -- An Auditor's Viewpoint," First National Symposium on the Management of Data Elements in Information Processing, January 1974, pp. 29-36.

[10] Collard, Albert F., "A Data Dictionary Directory," Journal of Systems Management, June 1974, pp. 22-25.

[11] Crawford, Perry, Jr., "On the Connection Between Data and Things in the Real World," First National Symposium on the Management of Data Elements in Information Processing, January 1974, pp.51-57.

[12] Crowe, Tom and J. H. Jones, "Fundamental Nature of Errors-- Data Capture; and the Use of a Data Dictionary," Management Datamatics, vol. 4 (1975) No. 3, pages 69-73.

[13] Cuozzo, D.E., and J. F. Kurtz, "Building a Base for Data Base: A Management Perspective," DATAMATION, October 1973, pp. 71-75.

[14] Curd, Edith F., "US Army Materiel Command (AMC) Progression From Reports Control To Data Element Management," First National Symposium on the Management of Data Elements in Information Processing, January 1974, pp.59-65.

[15] Curtice, Robert M., "Some Tools for Data Base Development," DATAMATION, July 1974, pp. 102-106.

[16] DATAPRO Directory of Software, DATAPRO Research Corporation, Delran, New Jersey.

[17] Date, C.J., An Introduction to Database Systems, Addison-Wesley Publishing Co., Philippines, 1975.

[18] Davis, Ruth M., "A Focus on the Role of the Data Manager," Second National Symposium on the Management of Data Elements in Information Processing, October 1975, pp. 57-60.

[19] Eberle, Sherron L., L. D. England, and Bernard H. Schiff, "The Use of a Data Base Management System for Standards Analysis," First National Symposium on the Management of Data Elements in Information Processing, January 1974, pp.67-77.

[20] Emswiler, C.E., Jr., and C. P. Heitzler, Jr., "The State's Model Motorist Data Base Project American National Standards Institute (ANSI) D-20," First National Symposium on the Management of Data Elements in Information Processing, January 1974, pp.79-102.

[21] FIPS PUB 28: "Standardization of Data Elements and Representation," December 5, 1973, 11 pages.

[22] Flynn, Robert L., "A Brief History of Data Base Management," DATAMATION, August 1974, pp. 71-77.

[23] Fong, Elizabeth, Joseph Collica and Beatrice Marron, "Six Data Base Management Systems: Feature Analysis and User Experiences," NBS Technical Note no. 887, November 1975.

[24] Gabrieli, E.R., "Semantic Coding and Data Element Characterization in Medical Computing," Second National Symposium on the Management of Data Elements in Information Processing, October 1975, pp. 83-89.

[25] Gomes da Costa, David, "Concept of Data Resource Management Vital to DBMS," Computerworld, September 24, 1975, p. S/12.

[26] GUIDE Report for the Data Base Administration Project: "The Data Administrator: Catalyst of Corporate Change," Draft Report, October 1, 1974, 53 pages.

[27] GUIDE Report for the Data Base Administration Project: "The Data Base Administrator," November 3, 1972, 76 pages.

[28] GUIDE Report for the GUIDE Dictionary/Directory Project: "Requirements for the Data Dictionary/Directory Within the GUIDE/SHARE Data Base Management System Concept," November 1, 1974. 77 pages.

[29] Hegland, Robert R., "The Standard Data Element System (STADES) for Controlling Data Elements Used in Navy Computer Programs for the Worldwide Military Command and Control System (WWMCCS)," First National Symposium on the Management of Data Elements in Information Processing, January 1974, pp.213-221.

[30] Hochman, Aaron, "Principles and Concepts of Data Resources' Management System Development," Second National Symposium on the Management of Data Elements in Information Processing, October 1975, pp. 91-98.

[31] Huffenberger, M.A., "The Design of Data Elements: A Data Base Perspective," Second National Symposium on the Management of Data Elements in Information Processing, October 1975, pp. 99-112.

[32] The ICP Software Directory, International Computer Programs, Inc., Indianapolis, Indiana.

[33] INPUT, "Analysis of the Market for a Generalized Data Dictionary," INPUT market survey report, Palo Alto, California, December 1975.

[34] Koch, H.S. and M.H. Gotterer, "An Independent Data Definition Facility for COBOL and FORTRAN," ACM-SIGFIDET Workshop, 1972, pp. 57-66.

[35] Lee, E. K. C. and E. Y. S. Lee, "Development of Data

Dictionary/Directory Using a Data Base Management System," Second National Symposium on the Management of Data Elements in Information Processing, October 1975, pp. 151-162.

[36] Lefkovits, Henry, Data Dictionary Systems, QED Information Sciences, Wellesley, MA, (To Be Published December 1976).

[37] Martin, George N., "Data Dictionary/Directory System," Journal of Systems Management, December 1973, pp. 12-19.

[38] McEwen, Hazel E., editor, Management of Data Elements in Information Processing, (Proceedings of a [First] Symposium Sponsored by the American National Standards Institute and by the National Bureau of Standards,) National Bureau of Standards, Gaithersburg, MD, NTIS No. COM 74-10700, January 1974, 474 pages.

[39] McEwen, Hazel E., editor, Management of Data Elements in Information Processing, (Proceedings of a Second Symposium Sponsored by the American National Standards Institute and by the National Bureau of Standards,) National Bureau of Standards, Gaithersburg, MD, NTIS No. PB 249-530, October 1975.

[40] McNamee, Patricia A., "Enforcing Naming Standards Through Use of a Data Dictionary," First National Symposium on the Management of Data Elements in Information Processing, January 1974, pp.263-270.

[41] Nerad, Richard A., "Data Administration as the Nerve Center of a Company's Computer Activity," Data Management, October 1973, pp. 26-31.

[42] Pan, George S., "The Characterization of Data Management Systems," Data Management, June 1971, pp. 18-23.

[43] Plagman, B. K., "An Integrated Corporate Data Base Concept and its Application," ACM-SIGFIDET Workshop, 1972, pp. 395-419.

[44] Plagman, B.K. and G. P. Altschuler, "A Data Dictionary/Directory System Within the Context of an Integrated Corporate Data Base," Fall Joint Computer Conference Proceedings, 1972, pp. 1133-1140.

[45] Pontius, James W., "The Need for Standardization of Data Elements and Data Codes-- From Origin of the Effort to Partial Fruition," First National Symposium on the Management of Data Elements in Information

Processing, January 1974, pp. 271-276.

[46] Puente, Fernando, "Systems Design Considerations For the US Army Materiel Command (AMC) Data Element Dictionary Directory System," Second National Symposium on the Management of Data Elements in Information Processing, October 1975, pp. 163-167.

[47] Rocke, Merle G., "The Need for Data Code Control," DATAMATION, September 1973, pp. 105-108.

[48] Roberts, John, "A Data Element Directory for a State Motor Vehicle Agency," Second National Symposium on the Management of Data Elements in Information Processing, October 1975, pp. 169-190.

[49] Shields, Curg, "An Integrated Dictionary for Systems and Data Components," Second National Symposium on the Management of Data Elements in Information Processing, October 1975, pp. 191-193.

[50] Sibley, E.H., and H.H. Sayani, "Data Element Dictionaries for the Information Interface," First National Symposium on the Management of Data Elements in Information Processing, January 1974, pp. 285-304.

[51] Simonette, Ilario, "Data Base Management System," Management Control, July/August 1975, pp. 102-107.

[52] Stern, Ludwig, "Contingency Planning: Why? How and How Much?" DATAMATION, September 1974, 6 pages.

[53] Taggart, William M. Jr., "An Information Documentation Language: A Framework for Deriving Information from Data," Second National Symposium on the Management of Data Elements in Information Processing, October 1975, pp. 195-210.

[54] Thomas, D. R., "DATAMANAGER -- A Freestanding Data Dictionary System," DATABASE Journal, vol.6, no. 5, pp. 14-17.

[55] Uhrowczik, P.P., "Data Dictionary/Directories," IBM Systems Journal, vol. 12, no. 4, 1973, pp. 332-350.

[56] U.S. General Accounting Office, "Emphasis Needed On Government's Efforts To Standardize Data Elements and Codes for Computer Systems," Report to the Congress, May 16, 1974, 69 pages.

[57] Walker, P. D., and S. D. Catalano, "Next in MIS: 'Data Managed' System Design," Computer Decisions,

December 1969, 4 pages.

[58] Wallis, Marvin G., "An Information Management View of Data Management," Second National Symposium on the Management of Data Elements in Information Processing, October 1975, pp. 221-230.

[59] Williams, Martha E., Scott E. Preece, and Sandra H. Rouse, "Data Element Analysis and Use of a Relational Data Base Structure for Mapping Bibliographic and Numeric Data Bases," Second National Symposium on the Management of Data Elements in Information Processing, October 1975, pp. 237-252.

[60] Wright, Arthur J., "Fundamental Tools and Techniques Toward Evolving a Data Element Management System," First National Symposium on the Management of Data Elements in Information Processing, January 1974, pp. 345-351.

[61] Young, Edith F., "Status of the Army Materiel Command's Progression from Reports Control to Data Element Management," Second National Symposium on the Management of Data Elements in Information Processing, October 1975, pp. 253-256.

NBS-114A (REV. 7-73)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBS SP-500-3	2. Gov't Accession No.	3. Recipient's Accession No.

4. TITLE AND SUBTITLE

Technical Profile of Seven Data Element Dictionary/Directory Systems

5. Publication Date

February 1977

6. Performing Organization Code

7. AUTHOR(S)

Belkis Leong-Hong and Beatrice Marron

8. Performing Organ. Report No.

9. PERFORMING ORGANIZATION NAME AND ADDRESS

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

10. Project/Task/Work Unit No.

11. Contract/Grant No.

12. Sponsoring Organization Name and Complete Address *(Street, City, State, ZIP)*

Same as item 9

13. Type of Report & Period Covered

14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

Library of Congress Catalog Card Number: 76-58915

16. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

A Data Element Dictionary/Directory (DED/D) is a software tool that is used to control and manage data elements in a uniform manner. It can serve data base administrators, systems analysts, software designers, and programmers by providing a central repository for information about data resources across organization and application lines. This report describes and classifies DED/D systems and discusses the potential benefits from their use. A technical profile of seven commercially-available DED/D systems is presented with side-by-side exposition of technical features.

17. KEY WORDS *(six to twelve entries; alphabetical order; capitalize only the first letter of the first key word unless a proper name; separated by semicolons)*

Computer software; data base management systems; data element dictionary; data element dictionary/directory; software tool.

18. AVAILABILITY

[X] Unlimited

[ ] For Official Distribution. Do Not Release to NTIS

[X] Order From Sup. of Doc., U.S. Government Printing Office Washington, D.C. 20402, SD Cat. No. C13.10:500-3

[ ] Order From National Technical Information Service (NTIS) Springfield, Virginia 22151

19. SECURITY CLASS (THIS REPORT)

UNCLASSIFIED

21. NO. OF PAGES

45

20. SECURITY CLASS (THIS PAGE)

UNCLASSIFIED

22. Price

$1.05

USCOMM-DC 29042-P74

ANNOUNCEMENT OF NEW PUBLICATIONS ON COMPUTER SCIENCE & TECHNOLOGY

Superintendent of Documents,
Government Printing Office,
Washington, D. C. 20402

Dear Sir:

Please add my name to the announcement list of new publications to be issued in the series: National Bureau of Standards Special Publication 500-.

Name _____

Company _____

Address _____

City _____ State _____ Zip Code _____

(Notification key N-503)

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH reports National Bureau of Standards research and development in physics, mathematics, and chemistry. It is published in two sections, available separately:

• Physics and Chemistry (Section A)
Papers of interest primarily to scientists working in these fields. This section covers a broad range of physical and chemical research, with major emphasis on standards of physical measurement, fundamental constants, and properties of matter. Issued six times a year. Annual subscription: Domestic, $17.00; Foreign, $21.25.

• Mathematical Sciences (Section B)
Studies and compilations designed mainly for the mathematician and theoretical physicist. Topics in mathematical statistics, theory of experiment design, numerical analysis, theoretical physics and chemistry, logical design and programming of computers and computer systems. Short numerical tables. Issued quarterly. Annual subscription: Domestic, $9.00; Foreign, $11.25.

DIMENSIONS/NBS (formerly Technical News Bulletin)—This monthly magazine is published to inform scientists, engineers, businessmen, industry, teachers, students, and consumers of the latest advances in science and technology, with primary emphasis on the work at NBS. The magazine highlights and reviews such issues as energy research, fire protection, building technology, metric conversion, pollution abatement, health and safety, and consumer product performance. In addition, it reports the results of Bureau programs in measurement standards and techniques, properties of matter and materials, engineering standards and services, instrumentation, and automatic data processing.

Annual subscription: Domestic, $9.45; Foreign, $11.85.

NONPERIODICALS

Monographs—Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

Handbooks—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications—Include proceedings of conferences sponsored by NBS, NBS annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

Applied Mathematics Series—Mathematical tables, manuals, and studies of special interest to physicists, engineers, chemists, biologists, mathematicians, computer programmers, and others engaged in scientific and technical work.

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a world-wide program coordinated by NBS. Program under authority of National Standard Data Act (Public Law 90-396).

NOTE: At present the principal publication outlet for these data is the Journal of Physical and Chemical Reference Data (JPCRD) published quarterly for NBS by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements available from ACS, 1155 Sixteenth St. N.W., Wash. D. C. 20056.

Building Science Series—Disseminates technical information developed at the Bureau on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NBS under the sponsorship of other government agencies.

Voluntary Product Standards—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The purpose of the standards is to establish nationally recognized requirements for products, and to provide all concerned interests with a basis for common understanding of the characteristics of the products. NBS administers this program as a supplement to the activities of the private sector standardizing organizations.

Consumer Information Series—Practical information, based on NBS research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace.

Order **above** *NBS publications from: Superintendent of Documents, Government Printing Office, Washington, D.C. 20402.*

Order **following** *NBS publications—NBSIR's and FIPS from the National Technical Information Services, Springfield, Va. 22161.*

Federal Information Processing Standards Publications (FIPS PUBS)—Publications in this series collectively constitute the Federal Information Processing Standards Register. Register serves as the official source of information in the Federal Government regarding standards issued by NBS pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

NBS Interagency Reports (NBSIR)—A special series of interim or final reports on work performed by NBS for outside sponsors (both government and non-government). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Services (Springfield, Va. 22161) in paper copy or microfiche form.

BIBLIOGRAPHIC SUBSCRIPTION SERVICES

The following current-awareness and literature-survey bibliographies are issued periodically by the Bureau:

Cryogenic Data Center Current Awareness Service. A literature survey issued biweekly. Annual subscription: Domestic, $20.00; Foreign, $25.00.

Liquified Natural Gas. A literature survey issued quarterly. Annual subscription: $20.00.

Superconducting Devices and Materials. A literature survey issued quarterly. Annual subscription: $20.00. Send subscription orders and remittances for the preceding bibliographic services to National Bureau of Standards, Cryogenic Data Center (275.02) Boulder, Colorado 80302.

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Washington, D.C. 20234

OFFICIAL BUSINESS

Penalty for Private Use, $300

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF COMMERCE
COM-215

SPECIAL FOURTH-CLASS RATE
BOOK