

NATL INST. OF STAND & TECH

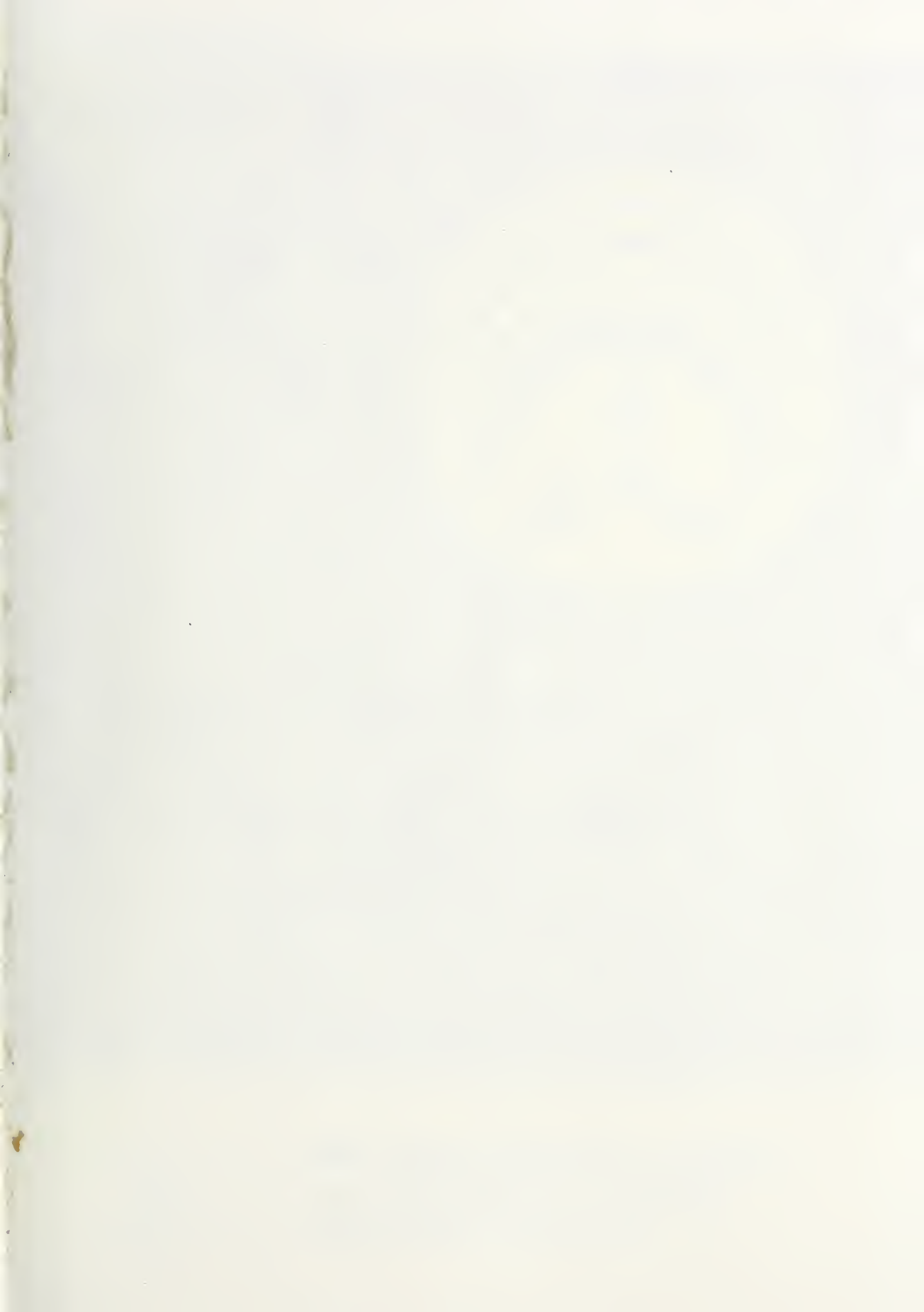


A11106 977962











A11103 090486

NATL INST OF STANDARDS & TECH R.I.C.



A11103090486

FIPS Task Group on D/Recommendations for  
QC100 .U57 NO.500-. 51, 1979 C.2 NBS-PUB

SCIENCE & TECHNOLOGY:

National Bureau of Standards  
Library, E-01 Admin. Bldg.

OCT 1 1981

191059

QC

100

.457



# RECOMMENDATIONS FOR DATABASE MANAGEMENT SYSTEM STANDARDS



**NBS Special Publication 500-51**  
**U.S. DEPARTMENT OF COMMERCE**  
**National Bureau of Standards**

## NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards<sup>1</sup> 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's technical work is performed by the National Measurement Laboratory, the National Engineering Laboratory, and the Institute for Computer Sciences and Technology.

**THE NATIONAL MEASUREMENT LABORATORY** provides the national system of physical and chemical and materials measurement; coordinates the system with measurement systems of other nations and furnishes essential services leading to accurate and uniform physical and chemical measurement throughout the Nation's scientific community, industry, and commerce; conducts materials research leading to improved methods of measurement, standards, and data on the properties of materials needed by industry, commerce, educational institutions, and Government; provides advisory and research services to other Government Agencies; develops, produces, and distributes Standard Reference Materials; and provides calibration services. The Laboratory consists of the following centers:

Absolute Physical Quantities<sup>2</sup> — Radiation Research — Thermodynamics and Molecular Science — Analytical Chemistry — Materials Science.

**THE NATIONAL ENGINEERING LABORATORY** provides technology and technical services to users in the public and private sectors to address national needs and to solve national problems in the public interest; conducts research in engineering and applied science in support of objectives in these efforts; builds and maintains competence in the necessary disciplines required to carry out this research and technical service; develops engineering data and measurement capabilities; provides engineering measurement traceability services; develops test methods and proposes engineering standards and code changes; develops and proposes new engineering practices; and develops and improves mechanisms to transfer results of its research to the ultimate user. The Laboratory consists of the following centers:

Applied Mathematics — Electronics and Electrical Engineering<sup>2</sup> — Mechanical Engineering and Process Technology<sup>2</sup> — Building Technology — Fire Research — Consumer Product Technology — Field Methods.

**THE INSTITUTE FOR COMPUTER SCIENCES AND TECHNOLOGY** conducts research and provides scientific and technical services to aid Federal Agencies in the selection, acquisition, application, and use of computer technology to improve effectiveness and economy in Government operations in accordance with Public Law 89-306 (40 U.S.C. 759), relevant Executive Orders, and other directives; carries out this mission by managing the Federal Information Processing Standards Program, developing Federal ADP standards guidelines, and managing Federal participation in ADP voluntary standardization activities; provides scientific and technological advisory services and assistance to Federal Agencies; and provides the technical foundation for computer-related policies of the Federal Government. The Institute consists of the following divisions:

Systems and Software — Computer Systems Engineering — Information Technology.

<sup>1</sup>Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D.C. 20234.

<sup>2</sup>Some divisions within the center are located at Boulder, Colorado, 80303.

JUL 25 1979

# COMPUTER SCIENCE & TECHNOLOGY: RECOMMENDATIONS FOR DATABASE MANAGEMENT SYSTEM STANDARDS

---

FIPS Task Group on Database Management System Standards

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



---

U.S. DEPARTMENT OF COMMERCE, Juanita M. Kreps, Secretary

Jordan J. Baruch, Assistant Secretary for Science and Technology  
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, Director

Issued August 1979

## **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 the series should complete and return the form at the end of this publication.

### **National Bureau of Standards Special Publication 500-51**

Nat. Bur. Stand. (U.S.) Spec. Publ. 500-51, 99 pages (Aug. 1979)

CODEN: XNBSAV

Library of Congress Catalog Card Number: 79-600087

U.S. GOVERNMENT PRINTING OFFICE

WASHINGTON: 1979

---

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402

Stock No. 003-003-02095-8 Price \$3.75

(Add 25 percent additional for other than U.S. mailing).

# TABLE OF CONTENTS

	Page
1. INTRODUCTION .....	1
1.1 BRIEF STATEMENT OF THE PROBLEM .....	1
1.2 BRIEF STATEMENT OF RECOMMENDED SOLUTIONS .....	2
1.3 PURPOSE AND METHOD OF FIPS TG-24 .....	4
1.4 ACTIONS OF OTHER STANDARDS BODIES .....	5
1.4.1 American National Standards Institute. ....	5
1.4.2 International Standards Organization. ....	6
1.4.3 Specification Work of Other Bodies. ....	6
1.4.4 Individual Federal Agency Standards Bodies. .	7
2. FEDERAL DBMS STANDARDS QUESTIONS .....	7
2.1 WHY DBMS STANDARDS? .....	7
2.1.1 Survey of DBMS Usage By Tg-24 Participants. .	7
2.1.2 Government DBMS Usage Trends. ....	8
2.1.3 Current Problems In Data Base Usage. ....	8
2.2 IS NOW THE TIME FOR DBMS STANDARDS? .....	8
2.2.1 Standards Impact On Database Technology. ...	8
2.2.2 Current Inventory of Standards Candidates. .	9
2.2.3 Sources of DBMS Standard Candidates. ....	9
2.2.4 Timeframe For Standards. ....	10
2.3 WHAT ARE THE ALTERNATIVES? .....	10
2.3.1 Alternatives To Any DBMS Standards. ....	10
2.3.2 Standards Other Than Federal. ....	11
2.4 FUTURE OF DBMS TECHNOLOGY .....	12
2.5 HOW ARE DBMS STANDARDS JUSTIFIED? .....	12
3. RECOMMENDATIONS FOR STANDARDS .....	13
3.1 APPROACH AND OVERVIEW .....	13
3.1.1 DBMS Components. ....	13
3.1.2 Structure of Recommendations. ....	13
3.1.3 General Assumptions. ....	13

3.1.4	General Benefits. ....	14
3.1.5	General Cost Considerations. ....	14
3.2	TERMINOLOGY .....	16
3.2.1	Background. ....	16
3.2.2	Recommendations. ....	16
3.2.3	Justification. ....	16
3.3	DATA DESCRIPTION LANGUAGE .....	17
3.3.1	Background. ....	17
3.3.2	Recommendations. ....	18
3.3.3	Justification. ....	19
3.4	DATA MANIPULATION LANGUAGE .....	21
3.4.1	Background. ....	21
3.4.2	Recommendations. ....	22
3.4.3	Justification. ....	22
3.5	DATA DICTIONARY/DIRECTORY FACILITY .....	26
3.5.1	Background. ....	26
3.5.2	Recommendations. ....	26
3.5.3	Justification. ....	27
3.6	QUERY LANGUAGE/END USER FACILITIES .....	28
3.6.1	Background. ....	28
3.6.2	Recommendations. ....	28
3.6.3	Justification. ....	28
4.	RECOMMENDATIONS FOR GUIDELINES .....	30
4.1	BACKGROUND .....	30
4.2	RECOMMENDATIONS .....	31
4.3	JUSTIFICATION .....	33
4.3.1	Assumptions. ....	33
4.3.2	Benefits and Cost Avoidance. ....	33
4.3.3	Costs. ....	33
5.	MISCELLANEOUS RECOMMENDATIONS .....	34
5.1	REDUCING STANDARD ADOPTION COSTS .....	34
5.1.1	Background. ....	34
5.1.2	Recommendations. ....	34
5.1.3	Justification. ....	34

5.2	DATABASE STANDARDS IN OTHER MAJOR AREAS .....	35
5.2.1	Background. ....	35
5.2.2	Recommendation. ....	35
5.2.3	Justification. ....	35
5.3	TRANSITIONAL STANDARDS ACTIONS .....	36
5.3.1	Background. ....	36
5.3.2	Recommendations. ....	36
5.4	PRIORITIES WITHIN THE RECOMMENDATIONS .....	37
5.4.1	Background. ....	37
5.4.2	Phasing Recommendations. ....	37
5.4.3	Justification. ....	38
6.	REFERENCES FOR TG-24 REPORT .....	39
7.	APPENDICES .....	43
7.1	APPENDIX 1 - FIPS TASK GROUP 24 CHARTER .....	43
7.2	APPENDIX 2 - DBMS USAGE SURVEY FOR TG 24 ....	44
7.3	APPENDIX 3 - DBMS TERMINOLOGY FOR TG-24 .....	56
7.4	APPENDIX 4 - DATA DESCRIPTION LANGUAGE .....	61
7.5	APPENDIX 5 - DATA MANIPULATION LANGUAGE .....	65
7.6	APPENDIX 6 - DATA DICTIONARY/DIRECTORY .....	80
7.7	APPENDIX 7 - END-USER/QUERY LANGUAGE .....	86

## PREFACE

Under Public Law 89-306 (Brooks Act), the National Bureau of Standards (Institute for Computer Sciences and Technology) has responsibilities to develop Federal Information Processing Standards (FIPS). Three major goals are sought in Federal standards: improved competition among vendors providing computer systems or services to the government, improved procurement procedures, and improved interchange of data and programs within the Federal government. FIPS publications may be either standards or guidelines. A standard is a precise statement of required functions or actions, while a guideline advises and suggests actions. Examples of standards range from the data codes for state abbreviations to the complex COBOL language specifications. Examples of guidelines include the published recommendations for physical computer security and privacy protection.

To assist the National Bureau of Standards in its consideration of FIPS standards, Task Groups are sometimes established to address specific subject areas. These Task Groups are advisory bodies made up of volunteer participants from Federal agencies. Task Group 24 on Data Base Management Systems is such a group. TG-24 purpose, scope, and program of work are contained in its charter, Appendix 1 of this report.

The issues addressed by Task Group 24 are important, complex, and highly technical. The recommendations of the Task Group are valued contributions to our work as representative statements of requirements. They are not necessarily the technical judgments or current positions of NBS. These views provide a concrete reference point for others to add their comments and recommendations. In each area addressed by the Task Group, NBS has underway a thorough study, including a cost-benefit analysis, leading towards a proposed Federal standard if warranted by the conclusions of continuing study. Our analyses coupled with continuing input from Federal agencies will guide the final decisions on Federal data base standardization. Consequently, we publish this report to invite additional comment. We will continue to seek comment as we proceed through the various steps toward standardization.

S. Jeffery, Director  
Center for Programming  
Science and Technology

## ACKNOWLEDGEMENTS

The National Bureau of Standards acknowledges its gratitude to the following TG-24 participants and their sponsoring organizations for their important contributions to the work of the Task Group.

Robert Ambrose, National Security Agency  
Christopher Avery, Department of Agriculture  
Ann Bandurski, Naval Ship Research Center  
John Berg, NBS, Chairman  
Edward M. Carter, USAF  
John Cooley, Federal Energy Administration  
Michael Corrigan, Defense Communications Agency  
Donald R. Deutsch, NBS  
Elizabeth Fong, NBS, Executive Secretary  
Harry Fujiwara, Department of Housing, Education and Welfare  
Dave Gurtner, Veterans Administration  
Rick Holt, National Library of Medicine  
Wassil Lagoey, US Army  
Lois Morin, Office of Management and Budget  
Howard Nathanson, Federal Home Loan Bank Board  
Janet Nathanson, Pension Benefit Guaranty Corp.  
Richard W. Pelc, General Services Administration  
Christy Pino, Central Intelligence Agency  
Oren P. Phipps, Defense Communication Agency  
Ralph Rutherford Jr., Defense Intelligence Agency  
Robert Schmid, US Army  
Ed Thomas, General Services Administration

In addition, others offered their ideas to the Task Group. These were:

Julian Adler, State Department  
Walter Frederic, Health, Education, and Welfare  
R. Michael Gall, Civil Service Commission  
Frank Manola, Naval Research Labs  
Paul Oliver, US Navy  
Neal Seago, Government Accounting Office  
Edgar Sibley, University of Maryland  
Dave Weinstein, Central Intelligence Agency  
Roxanne Williams, Department of Agriculture

## EXECUTIVE SUMMARY

The need for standardization in database management results from the increased usage of database software and the increased demand for data interchange within the Federal government. To meet this need, Task Group 24 recommends the following actions:

### ++Terminology.

1. A standard set of data base oriented terms should be established, should be coordinated with the work of ISO, and should be included in the current ANSI standard data processing glossary.
2. Guidelines should be established which encourage DBMS developers to use this glossary when describing database concepts.
3. Guidelines should be established whereby vendors are encouraged to utilize DBMS language syntax which is compatible with this glossary.

### ++Data Description.

1. A two-part data description standard is required that contains the common description of the data element (attributes) and the facility to describe multiple data structure classes.
2. The specification of the standard description of data attributes should be similar to the attributes in the PICTURE and TYPE clauses of the CODASYL DDL.
3. The specification of the standard description of data structures should be required to encompass current data models such as the hierarchical, network, and relational.
4. Consistent with recommendation 1, the standard description of the network data structure within the DDL should be based on the CODASYL DDL.
5. Consistent with recommendation 1, companion data structure descriptions for the hierarchical and relational data models should be developed.

## ++Data Manipulation.

1. Develop multiple Data Manipulation Language standards specifications.
2. [Develop a data manipulation language standard specification:]
  - (a) For each standard host programming language (e.g., COBOL, FORTRAN), develop immediately a standard DML specification for each category identified by TG-24.
  - (b) As a short range goal, develop a single standard DML specification for a given category that interfaces with all standard host programming languages.
  - (c) As a long range goal, develop a single standard DML specification containing the functionality of all categories.

## ++Data Dictionary.

1. Data dictionaries used by Federal agencies must be able to produce the standard DDL attribute description as recommended in Section 3.3.2 by the DDL Subcommittee of TG-24. (TG-24 took no position on the standardization of data dictionaries but addressed only those data dictionaries with an interface between the data dictionaries and the DBMS which must be standardized.)
2. The design of data dictionary must be capable of combining standard data attribute descriptions with data structure descriptions to generate the DDL for one or more DBMS.
3. Establish guidelines on the data dictionary usage.

## ++End-user/query Facilities.

1. Standardization of syntax and semantics of end-user facilities is not required. Such facilities are easily learned, problem and subject-matter dependent, and there are a diversity of end-user facility "styles."
2. Guidelines should be developed to aid the specification of requirements of end-user facilities for Federal procurement purposes.

++Standards Adoption.

1. The standards recommended should not be mandatory until other factors determine a change to the relevant systems.
2. Standards can be required for new applications or systems without requiring existing systems to also conform to the standard.

++Other Major Areas.

1. TG-24 recommends that the database standards actions described above apply to database facilities provided by computer services, distributed systems, and minicomputers.

RECOMMENDATIONS FOR  
DATABASE MANAGEMENT SYSTEM STANDARDS

The Federal Information Processing Standards  
Task Group on  
Database Management System Standards  
(FIPS TG-24)

In March, 1977 FIPS Task Group 24 initiated a study of the need for database standards within the Federal government. The voluntary participants from several Federal agencies considered the actions of other standards bodies; reviewed the alternatives to Federal standards; examined the issues of standards adoption, timing, and impact on technology; developed a method for justifying standards, and attempted to anticipate likely database technology advancements.

TG-24 recommended standards in certain specific technical areas, concluded that standards were premature in others, and emphasized the need for certain guidelines.

This final report of TG-24 contains the recommendations for standards and guidelines as well as the assumptions, benefits, and costs considerations used to justify the recommendations.

Key words: Database; DBMS; data-description; data-dictionary; data-directory; data-manipulation; languages; query; standards.

## 1. INTRODUCTION

### 1.1 BRIEF STATEMENT OF THE PROBLEM

Federal government usage of database technology, like the rapid growth noted in private industry, increases yearly. Each database system differs from the others and inhibits the interchange of skilled personnel, programs or data among these different systems. Even similar systems have slight differences that prevent quick interchange.

The resultant growth in the number of diverse database system undermines Federal goals sought by the standardization of programming languages and other software components. In view of the predicted hardware conversions within the Federal government, and the even greater likelihood of operating system and peripheral device changes over the next ten years, database technology will not meet its potential of facilitating conversion and may even worsen the conversion problem.

Current standards work is being performed in several areas of database technology but many other areas are being overlooked. While voluntary DBMS standards actions in the American National Standards Institute(ANSI) are underway, these actions require close, cooperative monitoring to insure that they will meet Federal needs and time frames.

## 1.2 BRIEF STATEMENT OF RECOMMENDED SOLUTIONS

This report contains a specific recommendation to develop a family of database standards for the Federal government. The Task Group recommends the following actions:

### ++Terminology.

1. A standard set of data base oriented terms should be established, should be coordinated with the work of ISO, and should be included in the current ANSI standard data processing glossary.
2. Guidelines should be established which encourage DBMS developers to use this glossary when describing database concepts.
3. Guidelines should be established whereby vendors are encouraged to utilize DBMS language syntax which is compatible with this glossary.

### ++Data Description.

1. A two-part data description standard is required that contains the common description of the data element (attributes) and the facility to describe multiple data structure classes.
2. The specification of the standard description of data attributes should be similar to the attributes in the PICTURE and TYPE clauses of the CODASYL DDL.

3. The specification of the standard description of data structures should be required to encompass current data models such as the hierarchical, network, and relational.
4. Consistent with recommendation 1, the standard description of the network data structure within the DDL should be based on the CODASYL DDL.
5. Consistent with recommendation 1, companion data structure descriptions for the hierarchical and relational data models should be developed.

#### ++Data Manipulation.

1. Develop multiple [Data Manipulation Language] standards specifications.
2. [Develop a data manipulation language standard specification:]
  - (a) For each standard host programming language (e.g., COBOL, FORTRAN), develop immediately a standard DML specification for each category identified by TG-24.
  - (b) As a short range goal, develop a single standard DML specification for a given category that interfaces with all standard host programming languages.
  - (c) As a long range goal, develop a single standard DML specification containing the functionality of all categories.

#### ++Data Dictionary.

1. Data dictionaries used by Federal agencies must be able to produce the standard DDL attribute description as recommended in Section 3.3.2 by the DDL Subcommittee of TG-24. (TG-24 took no position on the standardization of data dictionaries but addressed only those data dictionaries with an interface between the data dictionaries and the DBMS which must be standardized.)
2. The design of data dictionary must be capable of combining standard data attribute descriptions with data structure descriptions to generate the DDL for one or more DBMS.

3. Establish guidelines on the data dictionary usage.

++End-user/query Facilities.

1. Standardization of syntax and semantics of end-user facilities is not required. Such facilities are easily learned, problem and subject-matter dependent, and there are a diversity of end-user facility "styles."
2. Guidelines should be developed to aid the specification of requirements of end-user facilities for Federal procurement purposes.

++Standards Adoption.

1. The standards recommended should not be mandatory until other factors determine a change to the relevant systems.
2. Standards can be required for new applications or systems without requiring existing systems to also conform to the standard.

++Other Major Areas.

1. TG-24 recommends that the database standards actions described above apply to data base facilities provided by computer services, distributed systems, and minicomputers.

### 1.3 PURPOSE AND METHOD OF FIPS TG-24

The work of FIPS TG-24 followed its charter which appears in Appendix 1 of this report. This charter contains FIPS TG-24's purpose, scope and program of work. Nineteen Federal agencies contributed to this report over the course of one year. Several sub-Task Groups addressed specific subjects and reported their findings to the committee. Viewpoints from experts in the field, Federal government agency managers facing database decisions, and published material on database technology assisted the Task Group in meeting its goals. Subcommittees were formed to investigate and propose standards for each component. The final overall recommendations were reviewed so that each component's recommendations are consistent.

In addition to describing specific functional capabilities, each subcommittee considered and addressed issues such as:

1. A single standard or multiple standards.
2. To standardize at the syntax level or functional level.
3. The nature of the interface between the components and the DBMS as a whole software system, and the interface between components.
4. The impact of the proposed standards on the future DBMS technology.
5. The cost and benefit of the proposed standards within each component.
6. An assessment of the cost of implementing and maintaining the proposed standards.
7. A plan for the achievement of the proposed standards.

These issues are individually addressed in the "Justification" subsections within the Chapter "Recommendations for Standards."

#### 1.4 ACTIONS OF OTHER STANDARDS BODIES

1.4.1 American National Standards Institute. In autumn, 1972, the American National Standards Institute (ANSI) committee on Computers and Information Processing committee (X3) through its Standards Planning and Requirements Committee (SPARC) established a Study Group on Data Base Management Systems with a charter "to investigate the potential for standards." The Study Group issued an interim report in 1975 and a final report in July 1977. [ANSI 75, TSIC 77]

The final report contained neither specifications for a recommended standard nor recommendations for any action for standardization of any existing products or specifications. The report does contain a "framework" which can be used to consider future standards actions.

After accepting the report, SPARC initiated three pertinent actions: referred actions for subschema data description language specifications and data manipulation language specifications to the COBOL committee, referred actions for a Subschema data description language specification and data manipulation language specification to the FORTRAN committee, and initiated a committee for data description language specifications.

1.4.2 International Standards Organization. Within the International Standards Organization (ISO), Technical Committee 97/Study Committee 5/Working Group 3 on DBMS meets semiannually and has a very ambitious program of work. Its Scope of Work includes:

1. Define concepts for conceptual schema languages
2. Define or monitor definition of conceptual schema language
3. Develop a methodology for assessing proposals for conceptual schema languages.
4. Assess candidate proposals for conceptual schema languages
5. Define concepts for conceptual level end user facilities
6. Define conceptual level end user facilities
7. Take cognizance of and react to other data base developments as appropriate
8. Develop vocabulary for Data Base Management Systems

1.4.3 Specification Work of Other Bodies. The Conference on Data Systems Languages (CODASYL) is a voluntary body that developed the Common Business Oriented Language (COBOL) and guided that language's evolutionary development. CODASYL developed detailed language specifications for a FORTRAN Data Manipulation Language and Subschema Data Description Language, a COBOL Data Manipulation Language and Subschema Data Description Language, a host-language independent Data Description Language, and a draft Data Storage Description Language. The FORTRAN specifications were published in January 1977 and the latter three will be published in April 1978.

Though not a standards body the impact of CODASYL work on languages is well known, and their work in the DBMS areas has had significant impact already. The various specification developing committees of CODASYL meet at intervals varying from every six weeks to three or four months. Approximately 25 different vendors and users are represented in the developmental committees. Proposals for improvements to the specifications arrive from world-wide sources, including the European Computer Manufacturing Association, the International Federation of Information Processing Societies, and several vendor user bodies. CODASYL specifications will continue to evolve and will be considered as

standards candidates by ANSI. However, CODASYL is not a standards making body and CODASYL specifications can not become standards through CODASYL actions alone.

1.4.4 Individual Federal Agency Standards Bodies. Among the larger Federal agencies participating in TG-24, it was noted that several had their own standards-making function and some were currently considering DBMS standards. For example, the Department of Agriculture has already established a policy to use exclusively one proprietary DBMS for most applications. Similarly, the Army reviewed various options related to reducing the number of DBMS used by its components.

## 2. FEDERAL DBMS STANDARDS QUESTIONS

### 2.1 WHY DBMS STANDARDS?

2.1.1 Survey of DBMS Usage By Tg-24 Participants. TG-24 surveyed DBMS usage of those Federal agencies participating in TG-24 in order to understand to some degree the actual usage of DBMS within the Federal agencies. The survey was not a "rigorous" one but it is useful in indicating the likelihood of significant trends in the Federal agencies. Details of the survey and its findings are in Appendix 2.

The informal results support the assumption that Federal agency usage of database systems parallels the growth of such systems in private industry. Within the 14 agencies surveyed, 57 distinct DBMS were found. A few years ago, only large, special purpose database systems were reported and these were primarily in the Department of Defense. Therefore, TG-24 inferred a significant growth in Federal DBMS usage in recent years. Detailed determination of the quantities of program code and data now committed to DBMS systems awaits a more comprehensive and carefully developed survey.

Contributing to the difficulty of finding information on Federal use of DBMS is the lack of a central repository of such information. TG-24 lacked the resources to conduct a formal user survey. Even when data may exist it is not in a form that permits ready synthesis of the information needed. TG-24 recognized the value such information would have for Federal planning and encouraged the development of such statistics but, at the same time, understood the expense

which may be involved in determining them.

2.1.2 Government DBMS Usage Trends. In a copyrighted survey of DBMS usage at 360/370 sites [IDC 77], International Data Corporation found that "of 861 sites in sample, 312 users (36.2%) reported usage of a DBMS at yearend 1976, and the figure climbed to 433 sites [planned] by yearend 1978 (50.3%)." If the same percentages hold true for the Federal Government's 8649 computers (1975), 3131 sites now use DBMS and 4350 will in 1978.

2.1.3 Current Problems In Data Base Usage. Several Federal agency managers of data processing shared with TG-24 the issues that concerned them. These were identified as:

1. Need for a common functional requirements specification checklist to aid in the procurement of database systems.
2. Need for guidance on when to use what database system and how best to achieve its intended objectives.
3. Fear of a single, universal standard that prevents effective use of databases because of that agency's particular needs.
4. Need to identify and standardize subcomponents.
5. Need for assistance in the procurement process.
6. Need for assistance in converting to the standard.
7. Need for a quick, easy, low volume access to data.
8. Need for a standard that assists in reducing change and promotes user control of changes to product specifications rather than vendor control.

## 2.2 IS NOW THE TIME FOR DBMS STANDARDS?

2.2.1 Standards Impact On Database Technology. The Task Group considered specifically the possibility that standardization at this time might have a harmful effect on database technology. The result of this consideration was to note how difficult such a hypothesis was to prove or disprove. Further, it was not clear whether such a question was a proper consideration of the Task Group. Stated more directly, the recommendations of the Task Group require justification of cost savings or cost avoidance for the Federal government as a whole. Such considerations of future DBMS

technology impacts could properly be considered under potential or future costs but may be secondary to immediate costs or benefits.

TG-24 reviewed various scenarios in which "premature" database standards might inhibit new technology. The Task Group determined that difficulties arose immediately upon trying to select a metric for considering technology growth. Certainly subjective conclusions can be found everywhere. However, any such conclusions must consider three factors -- the type of standard concerned, the manner in which the selected standard is promulgated, and the extent to which it is accepted. Different combinations of these factors will affect different stages of program development cycle with differing impact on technology.

TG-24 looked for precedents in past standards activities. It reviewed the ASCII, COBOL, and MUMPS standards history and found no obvious instance where standards have inhibited or adversely affected technological growth.

2.2.2 Current Inventory of Standards Candidates. The size of the inventory of potential candidates for standardization depends on the degree of detail required for the statement or specification of the candidate. For example, the various CODASYL specifications are quite detailed statements of syntax and semantics presented in a very formal manner. On the other hand, some have argued that the user's manual of, say, a proprietary database system is equally a specification. Proprietary systems raise a second issue: the availability of the specifications for use by all interested implementors.

2.2.3 Sources of DBMS Standard Candidates. Identifiable sources of standard candidates do exist and should be examined. To aid its work, TG-24 reviewed the following sources:

- o Existing computer vendor software
- o Existing proprietary packages
- o Existing specifications from volunteer developmental groups
- o Existing Federally "owned" software
- o NBS developed specifications

In considering each of the broad areas, TG-24 noted a set of common requirements:

- o A clear and precise specification that permits any vendor to implement it
- o The general availability of such a specification
- o A concrete method to validate the specification and mediate disputes
- o A body that reviews and updates the specification
- o A method to validate implementations of the specification and mediate different interpretations of it
- o The need to deal with the impact on those systems not selected

2.2.4 Timeframe For Standards. Significant time requirements enter in the consideration of standards and especially so for DBMS standards. The Task Group concluded that a ten year life-cycle was the proper timeframe in which to consider DBMS standards. One to two years may be required for the development of the standard specifications. This time is not included in the ten year period. Another important time consideration is the time from the availability of the specifications to the availability of the first implementation. This period may also be one to two years. The ten year period will include periodic reviews.

Development of totally new DBMS specifications would require significantly longer periods before the preparation of specifications and the availability of useful products.

## 2.3. WHAT ARE THE ALTERNATIVES?

In deciding which approach to take in standardizing DBMS, consideration must be given to existing products, the feasibility of developing totally new products, and the work of other standards bodies. At the same time, the feasibility of implementation, the timeframe of the standard's development, the possible longevity of the standard, and the manner in which this rather large subject, DBMS, is divided into discrete components will also affect the decision.

2.3.1 Alternatives To Any DBMS Standards. Several alternatives to standards do exist:

1. Effective conversion tools would reduce the need for standards. However, "Data Base Directions II--the Conversion Problem" [BERG 78] reported the findings of a group of experts which stated that conversion technology has a need for standards in the area of

data interchange formats. Given such a standard, a translator that quickly and cheaply converted any database to any other database might eliminate the need for standards.

2. Data Dictionary/Directory standards may result in lessening the need for DBMS standards by providing common descriptions for all DBMS to use.
3. The Data Description Language used to provide, bridge, or interface with several Data Manipulation Languages may permit a multiplicity of DML's while providing many of the standard's objectives.
4. For smaller agencies, the availability of a single time-sharing DBMS may be used to satisfy the standardization objectives for Federal applications requiring the interchange of data and programs.

2.3.2 Standards Other Than Federal. The Federal government could adopt standards from other bodies or take advantage of existing de facto standards. These include:

- o ANSI Standards
- o De Facto Vendor Standards -- in the absence of other standards, the vendor's practice of making the same product available to all of its customers simplifies its task of maintaining and correcting the software it supports. This leads to a general compatibility between users of the same vendor systems. Generally, users groups have developed to exploit this compatibility through the interchange of information among the users. However, such de facto vendor standards are under the control of the vendor and subject to changes needed to meet vendor goals. While vendors are sensitive to customer needs, experience has shown that such changes have occurred.
- o Proprietary system issues -- Owners of proprietary systems may not wish to give up ownership or be unable to provide precise specifications of existing systems. At the same time, the competitive nature of the computer industry may result in a gradual development of proprietary systems toward the standard [BROC 76].
- o International Standards

- o Individual Government Agency Standards
- o Functional Application Area Standards For example: Law enforcement, health, library, air transportation, etc.

## 2.4 FUTURE OF DBMS TECHNOLOGY

TG-24 concluded after a review of pertinent references that database technology would continue to be characterized by significant and important work. The Task Group agreed with those who see change as inevitable over the next ten years. However, the Task Group concluded that change would be evolutionary and assumes that no "breakthroughs" will probably occur in that time period. Further, TG-24 concluded that proper standards planning will enhance the ability of Federal agencies to cope with breakthroughs if they should occur.

## 2.5 HOW ARE DBMS STANDARDS JUSTIFIED?

The first step in any DBMS standardization effort is developing a proposal for a particular specification. The next step is to justify the commitment of resources to develop the specification. TG-24 used a combination of quantitative and qualitative analysis to justify its recommendations. The Task Group began by systematically identifying the costs and benefits of DBMS standardization and documenting the underlying assumptions. This work assisted TG-24 in developing a set of priorities that led to the consideration of an interrelated set of DBMS components which it identified as a "family of standards."

The work will also assist those developing the specifications to compare actual experience against the assumptions in order to guide the specification development effort.

The difficulty in finding quantifiable costs led the Task Group to a method depending primarily on qualitative assessments. Therefore, the justification in the Recommendation Section is presented in qualitative terms.

### 3. RECOMMENDATIONS FOR STANDARDS

#### 3.1 APPROACH AND OVERVIEW

3.1.1 DBMS Components. Dividing DBMS into functional modules which together would meet the diverse needs of the Federal agencies provided the basis for organizing the TG-24 standards investigation and recommendations. The components that TG-24 chose as requiring standards considerations are:

- o Data Dictionary/Directory Facilities
- o Data Description Facilities
- o Data Manipulation Facilities
- o Query and End-User Facilities

3.1.2 Structure of Recommendations. Each of the components considered is presented in a form indicated by the following outline.

- I. Background (problem statement, scope and definition  
major approach and methodology)
- II. Recommendations (stated briefly and clearly)
- III. Justifications
  - (a) Assumptions
  - (b) Benefits and Cost Avoidance
  - (c) Costs (standard implementation, maintenance  
and usage)
  - (d) Discussions

All technical details and figures will appear in the appendices.

3.1.3 General Assumptions. In listing the assumptions to justify each of its recommendations for each component, TG-24 found that some assumptions were common to all or to many of them. Therefore, a list of general assumptions that are applicable to the Federal DBMS standardization effort as a whole follow:

- o Life cycle of database standard is ten years.
- o The use of DBMS will continue to increase.
- o The databases will grow in size, number, and complexity within the next ten years.
- o Changes in hardware and software over the next ten years will have little effect on the basic functions performed by data base systems.
- o Present DBMS architectures will be useful even in emerging new system environments (such as distributed systems) and changes to the DBMS architectures will be evolutionary over the next ten years.
- o There will be more DBMS conversions within each of the Federal organizations.
- o Transfer of data among Federal agencies within legislated guidelines will be increasing. Cost and benefits considered will be limited to those of Federal agencies.

3.1.4 General Benefits. The general benefits that can be identified for DBMS standardization as a whole are as follows:

1. Easier data conversion.
2. Easier program conversion.
3. Improved personnel transferability.
4. Improved data sharing among different computer installations.
5. Improved DBMS competition among vendors.
6. Improved selection and evaluation of DBMS products
7. Improved DBMS procurement process through larger vendor choice and competition.

3.1.5 General Cost Considerations. The general cost considerations parallel the set of common standards requirements discussed in the section dealing with the purpose and approach of TG-24. These are the costs of:

1. Developing a standard specification. Such a specification can be obtained from volunteer organizations, a government agency, vendors releasing their proprietary rights, or from a body established to provide the specifications.
2. Publishing the specification and analyzing the comments received.
3. Validating the specifications and mediating the differing interpretations.
4. Validating the implementations for conformance to the specifications.
5. Maintaining specifications over standards' life span.

While some of these costs are one-time costs, several are on-going costs over the life cycle of the DBMS standard. In addition to the costs associated with developing the standard specification, some costs can be identified with installing the standard DBMS component.

Note that costs associated with converting to a standard can be avoided by requiring use of the standard only when another form of change forces a conversion. Then the conversion to the standard has no direct costs. For example, first time users selecting a data dictionary system would probably pay no more for a data dictionary that produces output compatible with standard database systems than they would for a non-standard dictionary system.

Similarly, costs resulting from the conversion to improved hardware or software capabilities may be timed to include changes to standard practices. Agencies may shift to standard practices in a piecemeal fashion as existing non-standard practices require significant changes.

Costs associated with installing a standard DBMS component when the above considerations are ignored include:

- o Personnel-training and temporary loss of productivity
- o Conversion of data
- o Translation costs of programs

Cost will vary with the degree of difference between the non-standard system and the standard.

## 3.2 TERMINOLOGY

3.2.1 Background. The Federal government currently uses many different data base management systems. Each of these systems is unique in that its developer has adopted terms to represent syntactic and semantic entities which fit his own environment. For this reason, many similar data management functions are described in terms which are quite dissimilar. In the same fashion, many terms which are shared between several DBMS may have quite different meanings.

TG-24 recognized that its members used DBMS terms in a dissimilar manner. Consequently, the terminology appearing in Appendix 3 of this report was developed to help TG-24 members in their written and oral communications. This terminology definition is not intended to be a standard.

Work to date by ANSI has established a standard glossary of data processing terms [ANSI 77]. The absence of data base terms, however, is quite pronounced. The International Standards Organization/Technical Committee 97 is currently involved in establishing such a glossary.

### 3.2.2 Recommendations.

1. A standard set of data base oriented terms should be established, should be coordinated with the work of ISO, and should be included in the current ANSI standard data processing glossary.
2. Guidelines should be established which encourage DBMS developers to use this glossary when describing database concepts.
3. Guidelines should be established whereby vendors are encouraged to utilize DBMS language syntax which is compatible with this glossary.

### 3.2.3 Justification.

#### ++Assumptions.

1. The proliferation of DBMS will foster invention of unique DBMS terms for similiar concepts.
2. Terminology standards will continue to be developed by ANSI and ISO.

++Benefits and Cost Avoidance. Benefits to be derived from establishing a common set of database oriented terms can be seen mainly in the areas of training and information

interchange. A terminology standard would permit preliminary training, (e.g. university, technical school) to address general knowledge of database concepts, these concepts being embodied in the database glossary. Training at more specific levels could therefore be shorter in duration and more productive.

In the same fashion, interchange of information between individuals using different database systems can be facilitated. Different features can be related to the glossary of terms, thereby providing a mapping of information about one system onto another.

++Costs. The cost involved in establishing and maintaining a terminology standard is relatively small since Federal agencies would rely on voluntary development. The pay-back for this effort would begin immediately and would easily justify the expense in establishing the standard.

### 3.3 DATA DESCRIPTION LANGUAGE

3.3.1 Background. The data description language (DDL) is defined as a stand-alone language that describes:

- o Attributes of data elements
- o Logical relationships among units of data (records, sets, etc.)
- o Logical structure of the database
- o Logical methods of access to data

The DDL does not include the definition of physical storage media.

The Federal Government currently uses a large number of database management systems. Often a single agency supports more than one DBMS in order to satisfy varying user needs. Each DBMS has its own language for describing the attributes and relationships of the stored data. This variety of data descriptions makes it impossible for agencies, or even different units within an agency, to easily interchange data. The people who are interested in using the data are forced to learn the specific DDL which describes it. When data is exchanged, a new description of data for the target system must be written and tested before the data can safely be used.

The current situation is both costly and time consuming. Therefore, the recommendations are aimed at reducing both the time and cost required to exchange data between different users.

### 3.3.2 Recommendations.

1. A two-part data description standard is required that contains the common description of the data element (attributes) and the facility to describe multiple data structure classes.
2. The specification of the standard description of data attributes should be similar to the attributes in the PICTURE and TYPE clauses of the CODASYL DDL.
3. The specification of the standard description of data structures should be required to encompass current data models such as the hierarchical, network, and relational.
4. Consistent with recommendation 1, the standard description of the network data structure within the DDL should be based on the CODASYL DDL.
5. Consistent with recommendation 1, companion data structure descriptions for the hierarchical and relational data models should be developed.

++Explanation. The separation of the description of data element attributes from the description of data structure has been recommended because a single standard for both will not serve the needs of the Federal community. Separation of data element descriptions from data structure descriptions allows maximum flexibility while benefiting from the advantages of standardization. It would provide a single standard for data element descriptions and multiple standards for data structure descriptions. The concept of separation has been advanced by such noted authorities in the field as E.F. Codd, M.E. Senko, and James Martin. [CODD 71, SENK 73, MART 77]

The attributes required to describe data elements by any DBMS are similar, although the syntax and semantics used are often very different. A single standard in this area would simplify the user's task in describing the data elements contained in the data base; it was felt that the choice of syntax used to describe data element attributes was irrelevant. The attributes of the CODASYL DDL were suggested because:

1. They are well documented with a formal specification.
2. They are supported by a maintenance group.
3. They are an extension of COBOL, a widely used computer language.
4. There are currently several DDL implementations based on the CODASYL proposal.

In the data structure area, both the characteristics and the semantics/syntax required to describe it are different. Three discrete data models have been identified, i.e., hierarchical, network, and relational. Each of these models has unique characteristics that require a unique structure description. A single standard would impose one of the above models on all Federal agencies regardless of user requirements. For example, a user with an ad-hoc retrieval requirement who may best be served by a relational data model would be needlessly constrained if a network were adopted as the single standard.

### 3.3.3 Justification.

#### ++Assumptions.

- o The amount of data description in the Federal Government will continue to grow.
- o The amount and complexity of data stored in database structures by the Federal Government will grow significantly.
- o The increase in DBMS usage will cause an increase in DDL training requirement.
- o The demand for interchanging of data between DBMS will grow.

++Benefit and Cost Avoidance. Savings will be mainly in the area of:

- o Personnel - After the initial investment of re-educating all personnel in the use of the standard, DBMS retraining costs would be minimized. Additional costs would be avoided by reducing cost of new training, and reducing the cost of low productivity coupled with high error rates during the period of developing expertise in the use of the new DBMS.

- o Transfer of Data Descriptions - Cost of transfer of DDL between source and target DBMS would be minimized because the DDL of the source DBMS could be directly compiled on the target DBMS. Opportunity will exist for more automated means of converting between different DBMS.
- o Transfer of Data - Transfer of data would be simplified because the use of the standard DDL would allow a two step conversion. The source data would be converted to an intermediate file. The intermediate file would be loaded on the target DBMS using the same DDL. Target users would have no difficulty in utilizing the data.

There are other equally important considerations that are not easily quantifiable. These are timeliness and quality of data. The disruption felt by end users during conversion is difficult to measure but is a real factor.

The standardization of DDL would encourage transfer of data between users. Currently, the time and cost involved in data transfer forces users to do without needed data, or to duplicate collection and maintenance of data existing elsewhere. The report of the Federal Paper Commission cited duplicate collection of data as a serious government-wide problem.

Standardization of DDL would also encourage wide use of DBMS in the Federal Government. Many agencies are hesitant to get "locked in" to a non-standard DBMS causing them to use far less efficient means of storing and accessing data.

Standardization of DDL, as recommended, would contribute both quantifiable and non-quantifiable cost savings in the Federal Government. In addition, efficient information processing methods would be fostered.

++Costs. The Federal government should encourage and participate in voluntary standard action. However, in order to effect the intended goals the Federal government should be prepared to accept the costs of developing and maintaining the Data Description Language specification. The costs will be initially the development of the specification, whether in voluntary standards groups or as a Federal effort, and the subsequent maintenance of the specification.

## 3.4 DATA MANIPULATION LANGUAGE

3.4.1 Background. Data Manipulation Language (DML) is the language which a programmer uses to cause data to be transferred between a program and the database. The DML may not be a complete language by itself. It may rely on a host programming language to provide the procedural capabilities required to manipulate data. A user application program is written using a mixture of host programming language statements and the DML commands. The DML provides the ability to interact with the database by giving commands to cause an action and by providing a means to receive responses from the database or the processors involved. The DML consists of several data manipulation commands or functions which may include the following: data retrieval, data addition, data modification, data deletion and modifications of data relationships.

The diversity of data manipulation functions and languages causes many problem areas which would benefit from standardization. The following problems occur with current DBMS technology when users must contend with two or more DBMS.

- o There exists a different (non-standard) DML for each DBMS.
- o The data manipulation functions of each DML differ in syntax and semantics.
- o The data manipulation functions also differ in syntax and semantics between the various host programming languages for a given DBMS. Specifically, the DML for the FORTRAN language interface is different from the COBOL language interface.
- o The functional capabilities provided by DML are different across the set of available DBMS. This causes problems in conversion and procurement.

TG-24 evaluated the DML of twelve of the more commonly used DBMS and discovered four groupings. Chart I in Appendix 5 shows the twelve DBMS and their DML. Charts A, B, C, and D show the four categories of DBMS identified, and examples of the DML for each category.

TG-24 developed the concept "category" to help it analyze potential standards. The term "category" is an ad-hoc concept used merely for the purposes of our analysis. Categories are groupings of DBMS based on some technical criteria. The criteria used by TG-24 to categorize DBMS were

the commonality of data manipulation functions and data structures. Categories should not be equated to data models, although data structures are used as a major criterion to categorize data base systems. The term "data model" refers to the type of data structuring permitted by a DBMS (e.g., network, hierarchical, relational, ..) and various data models appearing in the literatures may have only slight differences. See for example [MART 77] and [KERS 76] for a discussion of different data models.

Though the concept of "categories" has special utility for TG-24's analysis, the idea behind this method may be of use to those who will be performing the follow-on detailed work. A more comprehensive analysis to categorize DBMS may require additional criteria. TG-24 does not wish to restrict the methodology used to categorize DBMS, but the intent of the analysis should be to support the goal of standardization by eliminating smaller differences, combining similar functionality, and producing the smallest number of categories.

The following recommendations assume four categories of DBMS with at most one standard DML for each host programming language for each category of DBMS and a potential of one standard DML for all DBMS.

#### 3.4.2 Recommendations.

1. Develop multiple DML standards specifications,
2. (a) For each standard host programming language (e.g., COBOL, FORTRAN), develop immediately a standard DML specification for each category identified by TG-24.  
  
(b) As a short range goal, develop a single standard DML specification for a given category that interfaces with all standard host programming languages.  
  
(c) As a long range goal, develop a single standard DML specification containing the functionality of all categories.

#### 3.4.3 Justification.

##### ++Assumptions.

1. Database management systems fall into limited identifiable categories reflecting different data structures.

2. Increasing numbers of application programs will be written using the DML.
3. The present interrelationships between data definition languages and data manipulation languages will remain the same in the next ten years.
4. The existing differences present in the host programming language interfaces are not necessary to provide the needed data manipulation functionality.
5. The increasing number of computer conversions anticipated within the Federal government will require reprogramming a large number of application programs using DML.
6. Data manipulation functions and data structures are valid criteria for categorizing DBMS.

#### ++Benefits and Cost Avoidance.

- o A greater return from (or a reduction of) required resources for programmer training, program conversion, data transfer, and application system transfer will occur with standard DML.
- o Limiting the number of standard syntactic and semantic definitions for data manipulation functions will simplify the terminology differences and make it possible to evaluate the capabilities of DBMS.
- o The evaluation, selection, and procurement of DBMS will be simplified by the standardization of a limited number of data manipulation languages.
- o The proposed standard DML interface will simplify data sharing in computer networks, community data bases, and distributed databases.

#### ++Costs. The following factors affect the standard's cost:

- o Increasing the number of categories and the number of host programming language interfaces will significantly increase the cost of specifying and developing the family of standard DML.
- o Conversion cost. A cost will be associated with converting existing DML to standard DML. The cost will vary with the differences between existing DML and the appropriate standard.

However, conversion costs can be avoided by:

- o installing the standard only for new applications.
- o delaying adoption of the standard until conversion is required by other factors.

++Discussion. The first of the recommendations posed above allows for the development of multiple DML standards. TG-24 seeks to permit flexibility in DBMS standards by providing a family of standards. Such an approach will allow an overall database architecture to exist that will provide useful functions and data structures that are not currently provided by any single DBMS using a single DML standard.

The second recommendation is a gradual movement towards a desirable end goal. It would be very beneficial to the DBMS end-user community if one standard DML was possible and practical. TG-24 felt that the standardization effort could not begin with the single DML standard route, but the effort may evolve to that end. There are major differences between the data structures and DML provided by various DBMS today. For this reason, if the DBMS continue to evolve, it is very possible that the major differences between DBMS will disappear and all DBMS will provide similar major capabilities. This would allow a single DML to be practical.

CATEGORY	HOST PROGRAMMING LANGUAGE					All
	COBOL	FORTRAN	PL/1	other		
Network						
Hierarchy						
Relational						
All						DM Function Superset

Figure 1 - Relationship of Host Language Standards to DML standards

The matrix in figure 1 illustrates the DML recommendations. Each of the categories listed down the left hand edge of the matrix provides the user with a useful viewpoint of the database. Across the top of the matrix appear examples of existing standard languages. As indicated by the matrix, each intersection of row and column provides an opportunity for a specific DML. For example, the CODASYL database specifications would be a candidate for the intersection of COBOL and the network category but, of course, would not satisfy the COBOL/relational intersection. Similarly, CODASYL has proposed a FORTRAN DML which is a candidate for filling the network category and FORTRAN intersection.

The DML recommendation identified each of the intersections as a potential standard. However, TG-24 also noted the inherent commonalities that would exist in all DML for any particular category. Save for differences of syntax reflecting the specific host language, the functions performed essentially remain the same over all the languages. One DML independent of any particular language could satisfy all the host languages for any particular category. The matrix indicates this with final right hand column which contains a DML for each category.

Finally, TG-24 considered the fact that all the different categories were treating the same important data. The concept of a superset of data manipulation functions common to all the categories but bundled to eliminate differences of syntax, data structures, or a particular viewpoint is not new. Several technical papers have treated the subject of "reconciling" various data models into a common language for the user. This lead to the conceptual possibility of "summing" the host language independent DML in the right hand column into the one common DML found at the bottom of the column.

The process of abstracting the various DML into a category-common DML and then, further, into a single, common DML is admittedly speculative and requires demonstrations of feasibility. However, such an approach would support directly the goals of standardization and TG-24 recommended its investigation.

### 3.5 DATA DICTIONARY/DIRECTORY FACILITY

3.5.1 Background. A data element dictionary/directory (to be referred to in short as data dictionary (DD)) is a software tool that is used to identify and interrelate data elements within an application or enterprise. It is viewed as the central repository of all descriptive information about each data element contained within an application database.

The scope of the standardization recommendations excludes data dictionaries that are manual tools for data resource management. In Appendix 6, the various types of data dictionaries, the various features that a typical data dictionary would provide, and some commercial package names are mentioned for illustrative purposes.

#### 3.5.2 Recommendations.

1. Data dictionaries used by Federal agencies must be able to produce the standard DDL attribute description as recommended in Section 3.3.2 by the DDL Subcommittee of TG-24. (TG-24 took no position on the standardization of data dictionaries but addressed only those data dictionaries with an interface between the data dictionaries and the DBMS which must be standardized.)
2. The design of data dictionary must be capable of combining standard data attribute descriptions with data structure descriptions to generate the DDL for one or more DBMS.

3. Establish guidelines on the data dictionary usage.

### 3.5.3 Justification.

#### ++Assumptions.

- o As databases continue to grow, the need to control data elements descriptions will increase.
- o Data dictionary usage in the Federal agencies will increase at a greater rate than the usage of DBMS within the next 10 years.
- o DD will be a critically important tool used by data base administrators for the central control of data-bases.
- o Distributed database management will enhance the need of data dictionaries.

#### ++Benefits and Cost Avoidance.

- o A data dictionary which produces standard DDL will reduce the cost of converting to the standard DDL.
- o The standard DDL produced by the DD to be interfaced to DBMS will reduce the need for manual coding which in turn will reduce errors.
- o There will be significant cost savings for transporting data for data interchange and for conversion purposes because a standard DDL would be generated from the data dictionary.
- o Data dictionaries will provide bridges to multiple DBMS.

++Cost. The cost of implementing this recommendation will be no more than the cost of implementing a data dictionary that does not produce a non-standard DDL.

++Discussions. Note that the recommendation proposed for the data dictionary is to standardize the interface between the DD and the DBMS. Compliance with this standard will permit the users of DBMS to use any data dictionary that has implemented the standard interface at no further cost.

## 3.6 QUERY LANGUAGE/END USER FACILITIES

### 3.6.1 Background.

There currently exists a tremendous variety of end user facilities for DBMS, and a substantial variety of taxonomies for these facilities. [LOUG 77]. Since many of these facilities are quickly learned, and many of them are designed for ad hoc activity, standardization of syntax and semantics appears unwarranted. Substantial costs may be incurred by a poor match between the facilities needed, and those provided by a procured DBMS. To avoid this, a match should be achieved prior to procurement.

### 3.6.2 Recommendations.

1. Standardization of syntax and semantics of end-user facilities is not required. Such facilities are easily learned, problem and subject-matter dependent, and there are a diversity of end-user facility "styles".
2. Guidelines should be developed to aid the specification of requirements of end-user facilities for Federal procurement purposes.

### 3.6.3 Justification.

#### ++Assumptions.

1. End-user facilities will be the subject of intensive investigation and rapid technical development, e.g., the role of intelligent terminals, is just beginning to emerge.
2. A variety of end-user facilities will be needed to accommodate the different needs and user populations.
3. The problem of matching end-user facilities to requirements will continue to be difficult.
4. Programming effort will continue to be expanded to improve end-user facilities where facilities are inadequate or inconvenient for the user group.

## ++Benefits and Cost Avoidance.

1. The major benefit of the procurement guideline would be improved utility of DBMS resulting from procurement of end-user facilities which best fit the needs and abilities of all user groups which are to use them.
2. The major cost avoidance resulting from the guidelines is due to the decreased need for "customizing" the end-user facilities, e.g., by programming new facilities using host languages and the data manipulation language, or by enhancing existing facilities.
3. Usage of the guideline will aid the user primarily during the procurement process. While not even an approximate estimate of the number of likely DBMS procurements can be made without a survey, one procurement error is likely to waste easily several months time. Additional uses of the guideline would be to guarantee retention of all existing functions during conversions, to provide guidance in upgrading existing facilities, and to assist database administrators in providing the appropriate tools to different user groups.

## ++Costs.

- o Development of the guideline should be relatively inexpensive. Maximum use should be made of existing studies and taxonomies of end-user facilities, and also of past procurement efforts where lists of specific requirements were used. A major part of the development effort should be a follow-up on the major procurements to check areas where experience after procurement indicates facilities were lacking.

## ++Discussion.

- o General comments. There are many examples of multiple DBMS in use by single organizations. A plausible explanation for this phenomenon is that the different DBMS provide different sets of end-user facilities which are so attractive to different user groups that the costs of multiple DBMS are outweighed by user convenience. It may well be that multiple end-user facilities are in fact required. It does not follow, of course, that multiple data dictionaries, data definition languages, and data manipulation functions are required - ideally different end-user facilities would make use of common system parts. The guideline should therefore clearly indicate that procurement officers may best

satisfy their requirements by purchasing several end-user facilities, from one or several vendors, plus possibly developing in-house facilities if appropriate, all of which use the same interface to other DBMS facilities.

- o Comments on first and second assumptions: A necessary and sufficient set of end-user facilities may eventually be defined as a result of current research. At that point, a standard for end-user facilities should be developed. These do not appear to be surveyable assumptions. The consensus of TG-24 is that the assumptions will remain valid for several years.
- o Comments on third and fourth assumptions: These assumptions have been informally verified by surveying the experience of TG-24 members. A formal survey might be appropriate to demonstrate more general validity.
- o Comments on end-user facility taxonomy: Three taxonomies of end-user facilities and user groups are presented in Appendix 7. They demonstrate possible approaches which might form the basis of a guideline. They also demonstrate the difficulties inherent in attempting to classify such a rich area. It is worth noting that when TG-24 was asked whether query facilities should contain update capabilities, one half was positive they should not, and the other half was positive they should.

#### 4. RECOMMENDATIONS FOR GUIDELINES

##### 4.1 BACKGROUND

Several areas or activities associated with database management systems do not easily lend themselves to strict standardization. These activities are concerned with such functions as the documentation of data base designs and implementations, ancillary operations performed in the database environment, and the establishment of database administration functions. This indicates the need for guidelines in addition to whatever standards may be necessary. These guidelines will treat subject matter that are, perhaps, premature for standardization, or offer too diverse a choice selection to permit hard and fast direction imposed

externally, or permit several equally good choices with no particular Federal-wide benefit achieved by pointing to a particular choice.

## 4.2 RECOMMENDATIONS

In addition to those guidelines recommended in the specific subject areas above, TG-24 concluded that guidelines recommending standards of good practice were needed in the following areas:

1. Computer security and privacy protection in database systems.
2. Procedures for database recovery, reorganization and audit. At this time, no Federal guidelines exist. A working panel report on database auditing [BERG 75] suggests the need for such a guideline. A group of audit experts [RUTH 77], supports this point, and discusses general audit administration, methodology and tools. While database recovery and reorganization procedures may be too DBMS-specific for inclusion into Federal guidelines, development of general guidelines or checklists may be possible.
3. Documentation - Currently FIPS Pub 38, "Guidelines for Documentation of Computer Programs and Automated Data Systems" contains guidelines for the description of database specifications. These guidelines describe the database specification for a database to be developed during the design stage of the development phase of the software life cycle. Guidelines will also be necessary for the documentation of the implemented database, i.e., after it is developed, tested, and loaded. This would be analogous to the separate guidelines that exist for the program specification and the program maintenance manual. The "Database Management Manual" might describe the database in its final, implemented form, and all common validation routines, recovery utilities, reorganization criteria and utilities, database administration support software, etc.
4. Performance monitoring - An important start has been made in this area with the issue of FIPS PUB 49, "Guideline on Computer Performance Management," in May 1977. In the context of database management,

guidelines could be developed for the monitoring of database growth and performance, general criteria for reorganization, database (re)design hints, etc.

5. DBMS evaluation and selection - FIPS guidelines on several activities in the analysis/evaluation/selection cycle would greatly assist Federal DP managers considering the database approach:
  - a. Analysis of existing ("conventional") data processing applications or potential applications (not yet automated) for implementation under database technology;
  - b. Methods for the comparison of available DBMS packages against application requirements and selection of best candidate(s);
  - c. Documentation practices for these activities,
  - d. Preparation of Request for Procurements - a need particularly noted by several managers that offered their comments to the Task Group, and;
  - e. Methods for benchmarking DBMS as a means to aid in evaluation of the responses to the Request for Procurement.
6. Database administration functions - Several existing publications from various sources could provide the basis for a FIPS guideline for the determination of the functional duties of Federal database administration staffs.
7. Requirements analysis - Little guidance now exists to assist Federal managers in preparing a good statement of their needs prior to seeking tools to meet these needs. Federal managers, particularly in the numerous smaller agencies, need such help.

## 4.3 JUSTIFICATION

### 4.3.1 Assumptions.

- o Useful guidelines can be compiled from existing material on good information practices.

### 4.3.2 Benefits and Cost Avoidance.

- o Guidelines can provide Federal agencies with approved practices drawn from industry experience and tailored to the special needs of Federal agencies. Examples of the special needs of Federal agencies include: procurement practices determined by regulations, statutory constraints of data processing action, and affirmative action programs with regard to information policy.
- o Guidelines can be maintained and modified to reflect Federal experience with them so that each agency can profit from the collective experience.

4.3.3 Costs. Costs associated with guidelines are the direct cost of compiling and reporting recommended standards of practice as well as that work needed to develop and test practices not available from industry sources. In addition, cost will be experienced in follow-up and verification of published guidelines.

## 5. MISCELLANEOUS RECOMMENDATIONS

### 5.1 REDUCING STANDARD ADOPTION COSTS

5.1.1 Background. To force the adoption of the standards recommended in this report universally and on a certain date could result in unnecessary expense to Federal agencies. Under such a requirement, conversion costs would occur in the training of personnel, the conversion of existing databases, and the translation of existing application programs. A less concrete cost would be the loss of productivity experienced by the agency during this period. These costs can be reduced and even eliminated by judicious timing of standard installation.

### 5.1.2 Recommendations.

1. The standards recommended should not be mandatory until other factors determine a change to the relevant systems.
2. Standards can be required for new applications or systems without requiring existing systems to also conform to the standard.

### 5.1.3 Justification.

#### ++Assumptions.

1. During a conversion to new systems requiring changes to programs and data, the use of standard products adds no additional costs to the conversion process.
2. New applications can be written in the standard product at little or no additional costs. Costs associated with maintaining two concurrent systems will be low and offset by reduction in conversion costs to move to the standard system when other system changes result in rewriting the existing programs

and converting the existing databases.

3. Incremental installation of standards will be feasible.

++Benefits and Cost Avoidance.

These recommendations will reduce conversion costs.

## 5.2 DATABASE STANDARDS IN OTHER MAJOR AREAS

5.2.1 Background. TG-24 considered certain other areas where database systems were major considerations. These were database systems provided by computer processing services (whether in batch or in the time sharing environment), distributed data processing or distributed data bases, and the mini/micro computer. TG-24 treated these issues only to the extent that it could conclude that the basic recommendations would apply to these systems as well.

5.2.2 Recommendation. TG-24 recommends that the database standards actions described above apply to data base facilities provided by computer services, distributed systems, and minicomputers.

### 5.2.3 Justification.

++Assumptions. The assumptions for these recommendations flow from the recommendations of each of the detailed standard recommendations. TG-24 assumes no significant differences will be required for each of the subject areas.

++Benefits and Cost Avoidance. Applying the same standards described previously to the database systems provided by computer processing services, to distributed systems, or to minicomputers will enhance the database work done in any of the other areas. It will insure the broadest possible sharing of people, programs, and data. Expanding the area of standards application will provide a greater payback for the costs expended in developing standards. Minicomputers, particularly, will benefit from the central discipline exerted by standards in view of their anticipated growth, diversity of vendors, and the lack of a common operating system discipline. Distributed systems, especially combining heterogeneous systems, will still have the conversion problem

between different systems but eased somewhat by the common standards shared by all the different systems.

++Costs. Cost associated with standard systems will not be different than that of non-standard systems. Conversion costs can be reduced by proper timing of standards usage.

### 5.3 TRANSITIONAL STANDARDS ACTIONS

5.3.1 Background. TG-24 recognized the timeframe of the many actions implied by its recommendations and recommends actions for transitional periods. Development of standard specifications may take as long as two years with another year after that for the implementation of a standard product. Standard products will remain merely an assertion of the vendor until validation procedures are available. What actions should Federal agencies take until a fully validated standard product is available?

The question suggests the existence of three time periods:

1. Until a specification is published.
2. From publication of the specification to its commercial availability.
3. From commercial availability to validation.

These periods of time will hold for each specification resulting from these recommendations. For example, the network DML and relational DML will have two distinct development paths which need not necessarily be concurrent. In addition, vendors may choose to combine step 2 and 3 by withholding a product until it has been validated.

5.3.2 Recommendations. TG-24 recommends the following positions for Federal agencies:

1. Prior to the development of standard specifications for a Data Description Language, Federal agencies can prepare for standardization within their own organization by using systems having a Data

Description Language similar to the CODASYL language.

2. In time period 1, select DBMS systems having independent Data Description Languages which describe data attributes similar to the CODASYL specifications.
3. In time period 2, require database systems under procurement to meet the standard specification of the Data Description Language and those standard Data Manipulation Language specifications which are appropriate for its application. For existing database systems, solicit translators from the existing system to the standard systems. Document existing systems using the standard specifications. Require Data Dictionary Systems and End User Facilities to meet the standard interface requirements.
4. In time period 3, require all new applications and procurements to use the standard systems.

#### 5.4 PRIORITIES WITHIN THE RECOMMENDATIONS

5.4.1 Background. Technical interrelations of the recommended family of database standards and payback considerations argue for establishing priorities in the development of the recommended standards. These priorities are reflected by the phasing recommended below.

##### 5.4.2 Phasing Recommendations.

++Phase 1. Develop specification for the common attribute part of the two part data description language standard as described in Section 3.3.2.

++Phase 2. Develop specification of the structural description for data structure classes as described in Section 3.3.3.

++Phase 3.

- o Develop specification for each data manipulation language.
  
- o Develop guidelines for using the data description language specification.
  
- o Specify interface requirements for data dictionaries and end-user facilities.

5.4.3 Justification.

++Discussion. In preparing these phasing recommendations TG24 notes that the network category has progressed further through these phases than the others. However, the recommendations are stated to avoid implying a preference for, or the inherent superiority of, the network category.

## 6. REFERENCES FOR TG-24 REPORT

### GENERAL

- [BROC 74] Brock, Gerald W., "The U.S. Computer Industry - A Study of Market Power" Ballinger Publishing Company, Cambridge, Mass. 1974.
- [CODA 76] CODASYL System Committee, "Selection and Acquisition of DBMS," available from ACM, 1133 Avenue of the Americas, New York, New York, 10036, March 1976.
- [CODD 71] Codd, E.F., "A Data Base Sublanguage Founded on the Relational Calculus," IBM Research Report, July, 1971, pp. 48.
- [CODD 71b] Codd, E.F., "Further Normalization of the Data Base Relational Model," IBM Research Report, August 1971, pp.33.
- [CODD 72] Codd, E.F., "Relational Completeness of Data Base Sublanguage," IBM Research Report, March 1972, pp.36.
- [COMM 77] The Commission on Federal Paper Work, "The Federal Information Locator Systems," July 1977, U.S. Government Printing Office.
- [DATE 75] Date, C. J., An Introduction to Data Base Systems, Addison Wesley Publishing Company, 1975.
- [GAO 77] General Accounting Office, "Millions in Savings Possible in Converting Programs From One Computer to Another," Report to the Congress, Sept 15, 1977.
- [IDC 77] International Data Corp., "360/370 Migration - 1977" A research report prepared by International Data Corporation, 214 Third Avenue, Waltham, Mass. 02154, July 1977.
- [MART 76] Martin, James, Principles of Data-Base Management, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1976.
- [MART 77] Martin, James, Computer Data-Base Organizations, 2nd Edition, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1977.
- [RUTH 77] Ruthberg, Z. G. & R. G. McKenzie (Eds.) "Audit and Evaluation of Computer Security," National Bureau of

- [SENK 73] Senko, M.E., E.B. Altman, et al., "Data Structures and Accessing in Data-Base Systems," IBM System Journal, Vol 12, No. 1, 1973, pp. 30-93.
- [SIBL 76] Sibley, Edgar H. "Computing Surveys: Special Issue on Database Management Systems" ACM, Computing Surveys, Vol. 8, No. 1, March 1976.

#### STANDARDS RELATED

- [ANSI 75] ANSI/X3/SPARC DBMS Study Group, "Interim Report on Data Base Management Systems," Feb. 1975.
- [BERG 75] Berg, John L. (Editor), "Data Base Directions - The Next Step" Proceedings of the Workshop of NBS and ACM held at Ft. Lauderdale, Florida, Oct. 29-31, 1975, National Bureau of Standards, ICST Special Publication No. 451.
- [BERG 78] Berg, John L. (Editor), "Data Base Directions - The Conversion Problems" to be published as National Bureau of Standards, ICST Special Publication, 1979.
- [CODA 69] CODASYL Data Base Task Group, "DBTG Report" Oct. 1969 (superseded by 1971 DBTG Report).
- [CODA 71] CODASYL Data Base Task Group, "DBTG Report" April 1971, Available from ACM, 1133 Avenue of the Americas, New York, N.Y. 10036.
- [SIBL 77] Sibley, Edgar H. "Standardization and Database Systems," IFSM TR No. 23, University of Maryland, also published in Proceedings of the Third International Conference on Very Large Data Bases, Oct. 1977.
- [TSIC 77] Tschritzis, D. & Klug, A. "ANSI/X3/SPARC DBMS Framework" University of Toronto, Computer Systems Research Group, Technical Note 12, July 1977.

#### COST BENEFIT

- [TG-5 77] Task Group 5 members, "Quantitative Methodology for Assessing the Economic Impact of Federal Information Processing Standards" Task Team Recommendations, March 21, 1977.
- [TG23 77] Task Group 23 members, "Preliminary Report of the Economic Analysis Study Group for MUMPS" Task Group 23,

July 26, 1977.

## TERMINOLOGY

[ANSI 77] American National Standards Committee X3 - Computers and Information Processing, "American National Dictionary for Information Processing" FIPS PUB 11-1, National Bureau of Standards, Sept. 1977.

## DATA DICTIONARY

[BRIT 77] British Computer Society, "Data Dictionary Systems Working Party Report," available from The British Computer Society, 29, Portland Place, London, W1N 4HU, United Kingdom, March 1977.

[CANN 74] Canning, Richard G., "The Data Dictionary/Directory Function" EDP Analyzer, vol 12, no. 11, November 1974.

[LEFK 77] Lefkovits, Henry C. "Data Dictionary Systems," Q.E.D. Information Sciences Inc. Wellesley, Mass., 02181, 1977.

[LEON 77] Leong-Hong, Belkis & Beatrice Marron, "Technical Profile of Seven Data Element Dictionary/Directory Systems" National Bureau of Standards Special Publication 500-3, February 1977.

[PLAG 72] Plagman, B. K. & 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.

[PLAG 77a] Plagman, B. K. "Data Dictionary/Directory System: A Tool for Data Administration and Control" Auerbach Information Management Series - Data Base Management, No. 22-01-02, Auerbach Publishers Inc., 6560 North Park Drive, Pennsauken, N. J. 08109, 1977.

[PLAG 77] Plagman, B. K. "Criteria for the Selection of Data Dictionary/Directory Systems," distributed at the Symposium on Management of Data Elements in Information Processing at National Bureau of Standards, September 1977.

[UHRO 73] Uhrowczik, P. P. "Data Dictionary/Directories," IBM Systems Journal, vol 12, no. 4 1973, pp. 332-350.

## DATA DESCRIPTION FACILITIES

[CODA 74] CODASYL Data Description Language Committee,

"CODASYL Data Description Language Journal of Development," National Bureau of Standards Handbook, 113, Jan. 1974.

[CODA 78] CODASYL Data Description Language Committee, "CODASYL Data Description Language Journal of Development," January 1978.

[CCA 76] Computer Corporation of America, "Data Descriptor Study," Final Report Contract DAAB-3-75-C-0464, August 16, 1976.

#### DATA MANIPULATION FACILITIES

[KERS 76] Kerschberg, L., Klug, A., & Tsichritzis, D., "A Taxonomy of Data Models." University of Toronto Technical Report CSRG-70, May 1976.

[THUR 77] Thurber, Kenneth J. & Patton, Peter J., "Data Structures and Computer Architecture" Lexington Books, DC Heath & Company, Lexington, Mass. copyright 1977.

#### END-USER FACILITIES

[LOUG 77] Lough, Dennis Elliot and Burns, Allen Dale, "An Analysis of Data Base Query Languages" NTIS ADA039783, March 1977.

## 7. APPENDICES

### 7.1 APPENDIX 1 - FIPS TASK GROUP 24 CHARTER

#### . DATABASE MANAGEMENT SYSTEMS (DBMS)

++Purpose. To make recommendations within one year to NBS and the FIPS Coordinating and Advisory Committee on the need for Federal database management system standards and any appropriate standards activities to meet such needs.

++Scope. The task group will consider all areas within current data base technology with a view towards defining relationships between DBMS and existing FIPS activities and a more precise statement of the scope for recommended standards activities. The area of study will include: distributed processing and databases, networking, data description languages, data manipulation languages, data dictionary/directory functions, DBMS support functions, and the role of mini-micro computers in DBMS.

#### ++Program of Work.

1. Determine the Federal need for DBMS standards. Consider Federal DBMS standards in such technical areas as DDL-DML, minicomputers and networking.
2. Survey existing DBMS models, determine the relative merits to the Federal agencies of the various models, and collect formal specifications.
3. Prioritize the various model specifications for detailed study.
4. For such model specification selected for detailed study, survey existing implementations and review to what extent each model specification feature was, in fact, implemented.
5. For each model specification selected for detailed study, make a formal recommendation for Federal action.

### BACKGROUND

An informal survey of DBMS usage was initiated at the first TG 24 meeting on March 15, 1977. All participants of TG 24 were asked to survey their own agency in the use of DBMS. The purpose of this informal survey was to gather preliminary data to substantiate the need for DBMS standards. In particular, the survey attempted to find out:

- o Whether Federal agencies are heavily committed to the use of DBMS.
- o If DBMS are used, how are they being used.
- o Whether TG 24 should conduct a formal survey on DBMS usage in order to assess the effect of a DBMS standard with respect to current operations.

The requested information for the informal survey follows:

- o Names of DBMS used within the agency. For each DBMS the following should be addressed:
  - o Type of application
  - o Type of usage
  - o Kind of user
  - o Type of support of central facility
  - o Data interchange

It was also agreed that a software system is a Data Base Management System if it possesses all of the following properties:

- o It is an integrated set of computer procedures
- o It facilitates usage of data
- o It facilitates storage and maintenance of large amounts of data
- o It provides frequently used shared functions
- o It potentially serves many functional purposes.

## SURVEY RESULTS

### Survey Population

14 Agencies responded with the DBMS usage survey. These are as follows:

1. Agriculture
2. Air Force
3. Army
4. Central Intelligence Agency
5. Defense Intelligence Agency
6. Federal Energy Administration
7. Federal Home Loan Bank Board
8. General Services Administration
9. Health, Education, and Welfare
10. National Bureau of Standards
11. National Security Agency
12. Office of Management and Budget
13. Pension Benefit Guarantee Corporation
14. Veterans Administration

Detailed data collected from each Agency is tabulated in Table 1. Summary data is presented below:

### Numbers of DBMS In Use Within Agencies

1 agency (FHLBB) has no DBMS at present.

1 agency (NSA) has 14 DBMS in use.

The remaining agencies have 1 or more DBMS. The conclusion from this fact is that most of the agencies are using DBMS and some agencies use more than one DBMS.

### In-House Written Versus Commercial Packages

Among 57 distinct DBMS in use in the survey population, 17 were in-house written. 6 were developed by the Air Force, 5 were developed by NSA. Roughly 75% of DBMS in use were commercial package.

### Characteristics of In-House Written DBMS

Two in-house written systems are used for document retrieval. The remaining in-house written DBMS are built for intelligence military applications. Real-time processing with some on-line and batch usage is typical of the in-house

written systems.

### Commercial DBMS Packages Distribution

(a) Host-Language Type systems - These include Codasyl-like systems but not necessarily Codasyl syntax. Systems include DMS-1100, TOTAL, DMS-II (Burroughs), DBMS-10, DBMS-11, IDMS. 7 agencies (Army, VA, Air Force, HEW/SSA, OMB, NBS, NSA) have these systems.

(b) System 2000 - 4 agencies (Army, Air Force, HEW/NIH, Agriculture) use this system.

(c) IMS - 2 agencies (Air Force, DHEW) use this system.

(d) Report Writers - 2 report writers (Mark IV, Easytrieve) are used by 2 agencies (Agriculture, NIH).

(e) Bibliographic systems - 4 bibliographic systems (CAIN, BASIS, 2 in-house written) are used among the survey population.

(f) Hardware supporting only one DBMS system - 2 DBMS can be classified as the sole DBMS operational for the hardware: IMAGE 3000 on Hewlett Packard, and DMS-II on Burroughs.

(g) Time-Sharing Service - 3 agencies (Army, GSA, Agriculture) use the CSC Infonet Time-sharing service which has System 2000, DML and Aladin. The Pension Benefit Guarantee Corporation uses Compu-Serv time-sharing service which has System 1022.

(h) Early generation DBMS systems - 5 early generation systems are used among the survey population. These are FFS, NIPS, IDS-I, MARS and MARK III. Users were system programmers and the usage is batch for report generation purposes.

(i) Research systems - 1 system INGRESS which is a relational DBMS developed by University of California, Berkeley, is used by 2 agencies (NBS, NSA). Usage is experimental research oriented and no production applications have been developed. The Army is using UNIBASE written in ANS Cobol 74 for research in DBMS portability.

### Usage and Users

(a) For Host-language and CODASYL-like systems, the usage is predominantly batch, transaction-oriented and used mostly to produce reports. Users were either subject-matter specialist or clerks for invoking a pre-defined transaction and system programmers for coding transactions.

(b) For systems such as System 2000, ADABAS, 1022, GIM-II, usage is predominantly on-line querying and updating.

(c) The majority of the users of DBMS are reported to be subject-matter specialists with programmers doing the system interface type of task.

(d) Agriculture has an experience with self-imposed standards. System 2000 at Agriculture has about 28 applications. Some are small applications with less than 100 records. The other DBMS in use at Agriculture are IMAGE-3000, Infonet DML and in-house written ones. Forestry uses FS-GIM. This might be due to the agency imposed standard DBMS practice to define all applications using one DBMS.

(e) In two cases (Army and NSA), DBMS-11 is being used experimentally as a back-end to a host computer 370/158.

TABLE 1 - DBMS USAGE SURVEY OF TG-24

Agency Name	DBMS and Hardware	Application Type	Usage and Users
Agriculture	System 2000 on IBM 370/168 and UNIVAC 1100	More than 25 various applications (some very small data bases).	Majority of usage is on-line ad-hoc retrieval and on-line updating; used by subject matter specialists.
	System 2000 (Infonet) on UNIVAC 1108	Official Estimate Data Base.	Batch weekly updates.
	IMAGE 3000 on HP 3000	Brucello's indemnity claims system.	On-line updates produce reports.
	FORESIGHT* on IBM 370/168	Budget Reporting System.	On-line updates.
	DML (Infonet) on UNIVAC 1108	Export Sales	On-line weekly updates.
	DML (Infonet) on UNIVAC 1108	Personnel	Batch bi-weekly updates.
	User-Writer* on IBM 370/168	Current research information.	Batch
	User-Writer (ISAM)* on IBM 370/168	Production, supply and distribution.	Batch
	FS-GIM on UNIVAC 1108	Range management systems.	Batch
	FS-GIM on UNIVAC 1108	Management attainment systems.	on-line
	FS-GIM on UNIVAC 1108	Timber management systems.	Batch
	CAIN on IBM 370/168	Bibliographic	Batch
	Easytrieve on IBM 370/168	GSA-ADO/MIS	Batch reporting

\* indicates in-house built DBMS

Agency Name	DBMS and Hardware	Application Type	Usage and Users
	MARK III on IBM 370/168	Federal Assistance Programs Retrieval Systems.	batch
	ALADIN (Infonet) on UNIVAC 1108	Administrative Record System	batch
Air Force	AFOLDS* (Air Force On-line Data System) on B-3500	Real property management, general accounting, hospital administration, form complex aircraft operational scheduling, project management, procurement, student scheduling, configuration analysis, library circulation control.	Transaction-oriented; used by subject matter specialist (non DP oriented).
	WWDMS* (World-Wide Data Management System), on Honeywell H-6000 series	Maintenance, force management, intelligence data collection (other) military application.	On-line query, updates are built through Honeywell's TSS; used by subject matter specialists.
	MIDMS* (Machine Independent DMS) on IBM 1410	Data Acquisition, real property management, real time data collection, intelligence data collection.	Ad-hoc querying; used by subject matter specialist.
	System 2000 on IBM 370	Munitions production scheduling, industrial funding, test data storage, project management, simulation data acquisition procurement.	Some on-line, through CICI and TP-2000, 65% batch, extensive report writer; used by system programmers and subject matter specialists.
	DMS-II (Data Management System--II) on B-6700	Personnel systems education.	Batch; used by students and subject specialists.
	FMIS* (Force Management Information Systems) on Honeywell 6000 series	Total Force Management (i.e., people, aircraft).	Real-time and batch, transaction-oriented; used by command and control specialists, subject matter specialists.

Agency Name	DBMS and Hardware	Application Type	Usage and Users
	VENUS* (Versatile, Expressive by Nature, Update/Query System) on IBM 370	Laboratory management, Large project control, student transcript, curriculum management.	On-line ad-hoc query and update; used by subject matter specialists.
	IMS (Information Management System) on IBM 370	Maintenance documentation.	On-line transactions with extensive batch reporting; used by aircraft mechanics/engineers, application programmers.
	FFS (Formatted File System) on IBM 360	Telemetry data collection and analysis.	Batch reduction of data where each usage requires new programming; used by experienced FORTRAN programmers.
	IDS-I (Integrated Data Store) on Honeywell 6000 series	Resource utilization accounting.	Batch reporting; used by support personnel and data automators.
	INFOCEN* (Information Central), developed by Mead Data Central on IBM 370	Over 150 various applications from project management through data acquisition.	On-line query/update; used by subject matter specialists.
Army	Aladin/Manage (Infonet) on UNIVAC 1108	Procurement, budget, safety and graphics applications.	On-line and batch usage.
	WWMCCS* on Honeywell 6000 Series	Mixed application types involving command and control functions.	On-line and batch usage.
	BASIS on CDC 6600	Historical data retrieval.	On-line usage.
	DMS-1100 on UNIVAC 1108	Budgeting system, personnel systems, research programs.	On-line and batch usage.
	System 2000 on UNIVAC 1108, IBM 370 and CDC 6600	R&D Program Management, personnel system, material and readiness, and inventory of dams.	On-line and batch usage.

Agency Name	DBMS and Hardware	Application Type	Usage and Users
	IDS I on Honeywell 437	Management information system.	Remote batch usage only.
	MARS on CDC 3300	Material and acquisition system.	On-line and batch usage.
	MIRADS on UNIVAC 1108	Testing only.	
	NIPS on IBM 360/40 and IBM 370/155	Military personnel, logistics, readiness applications.	On-line and batch usage.
	TOTAL on IBM 370	Material and acquisition, management information systems.	On-line and batch usage.
	NOMAD on IBM 370	Facilities management and graphics application.	On-line usage.
	IDMS on DEC 11 and IBM 370	Backend processor research and development and management information system.	On-line usage.
	UNIBASE on machine with ANS COBOL	Management information systems research and development.	On-line and Batch usage.
Central Intelligence Agency	GIM-II on IBM 360/370	Primarily for administrative data bases such as personnel, finance and budget, equipment and supply, inventory, contracts, project trackup, and analyst automation	Ninety percent used for on-line updating, report generation is batched; used by subject matter specialists and programmers for design and testing.
	NIPS on IBM 360/370	Personnel, inventory, operational support, document tracking, and project management.	Batch report generation and on-line users do statistical calculation; used by programmers and analysts.

Agency Name	DBMS and Hardware	Application Type	Usage and Users
	RAMIS on IBM 360/370	Tracking system; i.e., tracking hardware, tracking VM users, tracking VM problems.	Only on-line; used by programmers and small groups of subject matter specialists.
Defense Intelligence Agency	In-house built*	Intelligence data base.	On-line interactive response to ad-hoc queries; used by world-wide intelligence users.
Federal Energy Administration	INQUIRE on IBM 370	Preparation and maintenance of payroll and personnel records.	On-line and batch usage.
	ADABAS on IBM 370	Generating reports for scheduled cycles and for "on-the-spot" requirements from a vast data base.	On-line and batch usage.
Federal Home Loan Bank Board	No DBMS at present	Financial reports data base.	Reporting - need for aggregate data; used by economic analyst, examiners, accountants.
General Services Administration	DML (Infonet) on UNIVAC 1108	Building management system, equipment tracking system.	On-line and batch usage.
Health, Education, and Welfare	IMS on IBM 370	Payroll, accounting	All batch; used by computer specialists.
HEW, Center for Disease Control	Virtual System Personal Computing* on IBM 370	Programming	Programming aids, used by programmers and analysts.
HEW, Social Security Administration	TOTAL on IBM 370	Beneficiary Data Base Health Insurance - Billing.	Query, format query; used by subject matter specialists and field personnel.

Agency Name	DBMS and Hardware	Application Type	Usage and Users
HEW, Social Security Administration	DMS 1100 on UNIVAC 1108	Personnel, administrative, statistical survey correspondence and case control.	Batch-demand, query update, query ad-hoc; used by subject matter specialists and programmers.
HEW, National Institute of Health	IMS (INQUIRY)* on IBM 370	Financial, Inventory	On-line and batch, transaction-oriented. Used by clerical, manager, system programmer.
HEW, National Institute of Health	MARK IV, TOTAL, S2K on IBM 370		
HEW, Food and Drug Administration	Model 204 on IBM 370	Personnel, inventory, correspondence control, grants, scientific.	Batch reporting, ad-hoc query, on-line/batch updating. Transaction-oriented; used by clerical and programmers.
National Bureau of Standards	DMS-1100 on UNIVAC 1108	Solar data, marketing data.	Mostly reporting; used by subject matter specialists and system programmers. Interactive query available through QLP.
National Security Agency	MIRADS on UNIVAC 1108	Solar grant data.	On-line query and batch reporting.
	INGRESS on PDP-11	Research only.	
	BASIS-70 on IBM 370	Document Retrieval	On-line and batch usage.
	DBMS-10 on DEC PDP-10	Magnetic tape production management.	Transaction-oriented, ad-hoc queries.
	DBMS-11 on DEC PDP-11	Experimentation in backing up with IBM 370/158 as host.	On-line and batch; used by programmers, systems.

Agency Name	DBMS and Hardware	Application Type	Usage and Users
	DMS-1100 on UNIVAC 1108	Message processing	Transaction, batch, on-line; used by programmers.
	Dris (in-house built)* on Honeywell HIS-6060	Document retrieval	On-line and batch usage.
	IDMS on DEC PDP-11	Production scheduling and reporting, distribution and accounting equipment modification.	On-line and batch; used by subject matter specialists.
	INGRESS on DEC PDP-11	Research and development	On-line and batch; used by engineering programmers.
	IDS-2 on Honeywell HIS-6060	MIS for R&D task management.	On-line and batch usage.
	MODEL 204 on IBM 370	Personal, payroll and inventory control.	On-line, batch and transaction. Used by programmers and subject matter specialists.
	Solis (in-house)* on Burroughs 7700	Document Retrieval	Batch; used by programmers and subject matter specialists.
	Specol (in-house)* on IBM 370	Intelligence Processing	Remote batch; batch and used by programmers and subject matter specialists.
	1022 on DEC PDP-10	Graphics display and retrieval.	On-line and real time; used by programmers and subject matter specialists.
	TIDE/AUTOLINE* on UNIVAC 494	Message processing	Batch usage.
	TIPS/COINS* on UNIVAC 494	Storage and retrieval of technical information.	Remote batch; used by programmers and subject matter specialists.

Agency Name -----	DBMS and Hardware -----	Application Type -----	Usage and Users -----
Office of Management and Budget	TOTAL on IBM 370	The report clearance office information.	For file maintenance only. Retrieval and reporting uses ASI-SI and COBOL programs. Transaction implemented by programmers; used by subject matter specialists.
	Budget Preparation System* on IBM 370	This system supports budget of the U.S.	Major usage is reporting financial data; used by budget examiners.
	Budget Status System* on IBM 370	This system is designed to aid in analyzing budget requests and long-range budget forecasts.	Major usage is reporting.
Pension Benefit Guarantee Corporation	1022 (Compu-Serv time-sharing service) on DEC PDP-10	Financial and administrative application.	On-line single request; used by 1 senior programmer and 3 trainees.
Veterans Administration	TOTAL	Veteran Benefits M/R Maintenance.	Transaction-oriented; used by subject matter specialists (non-DP oriented).
	DM-IV (IDS/II)	Veteran Benefits M/R Maintenance.	Transaction-oriented and used by subject matter specialists (non-DP oriented).

### 7.3 APPENDIX 3 - DBMS TERMINOLOGY FOR TG-24

- ACCESS - The operation of seeking, reading, or writing data on a storage unit. [MART 75]
- ACCESS CONTROL - The mechanism in a data base system which provides protection of the data base from both intentional destruction and improper disclosure. [DATE 75]
- ACCESS METHOD - A technique for moving data between a computer and its peripheral devices, e.g. serial access, random access, remote access, virtual sequential access method (VSAM), hierarchical indexed sequential access method (HISAM). [MART 75]
- AUDIT TRAIL - The process of keeping of records in both events and data activities within a system to allow future examination and verification of what has transpired. [CODA 76]
- AUTHENTICATION - The process of supplying information known only to the person the user has claimed to be. [DATE 75]
- AUTHORIZATION - The definition to the system of the operations each user is allowed to perform. [DATE 75]
- BACK-UP - The process of generating a copy of a data base at some point in time. [CODA 76]
- CATEGORIES - Categories are groupings of database management systems based on some technical criteria.
- DATA ATTRIBUTE - A known characteristic of a data item. (e.g., a numeric field, a date field, etc.)
- DATABASE - A collection of interrelated data items processable by one or more programs.
- DATABASE ADMINISTRATOR (DBA) - A person or persons given the responsibility for the definition, organization, protection and efficiency of the database for an enterprise.
- DATA DICTIONARY (DD) - A software tool that is used to control the totality of data elements within an application. It is the central repository of all descriptive information about each data element.

DATA DESCRIPTION LANGUAGE (DDL) - The language used to describe the database, or that part of the database known to a program.

DATA ELEMENT - A named collection of one or more items.

DATA INDEPENDENCE - The property of being able to change the overall logical or physical structure of the data without changing the application program's view of the data. [MART 75]

DATA ITEM - The smallest unit of named data containing no sub-structure; the lowest level of addressable data in which data value(s) are physically stored.

DATABASE MANAGEMENT SYSTEM - A software system is a database management system if it possesses all of the following properties:

- . It is an integrated set of computer procedures
- . It facilitates usage of data
- . It facilitates storage and maintenance of large amounts of data
- . It provides frequently used shared functions
- . It potentially serves many functional purposes.

DATA MANIPULATION LANGUAGE (DML) - The language used to cause data to be transferred between the application program and the database.

DATA MODEL - See Data Structure.

DATABASE POPULATION - The process of placing occurrences of data values into a defined data base. [CODA 76]

DATA SECURITY - The protection of the data in the data base against unauthorized disclosure, alteration, or destruction. [DATE 75]

DATA STORAGE DESCRIPTION LANGUAGE (DSDL) - A language to define the representation of a data base in storage. This term is used in [CODA 78].

DATA STRUCTURE - The logical relationships which exist among the units of data in a database and which are under control of a database management system.

DATA VOLATILITY - Refers to the rate of change of the values of stored data. [CODA 76]

DEVICE MEDIA CONTROL LANGUAGE (DMCL) - A language used to map the data onto physical storage media. It describes the physical location and organization of the data.

DISTRIBUTED DATABASE - A database under the overall control of a central database management system, but where the devices on which it is stored are not all attached to the same processor.

END USER - The end users are persons who perform the application functions. End users include parametric users and generalized function users, but they are not system support personnel. [CODA 76]

EXTRACTION - The process of obtaining data values on each data structure level and identifying the disposition of these values to an output media. [CODA 76]

FILE - A collection of all occurrences of a given type of logical record.

HIERARCHY - A set of directed relationships between two or more units of data, such that some units are considered owners while others are members. This is distinguished from a network in that in a hierarchy, each member can have one and only one owner.

HOST LANGUAGE SYSTEM - A database management system that is built upon the facilities of a programming language and is identified to the application programmer for logical and physical file manipulations. The tools are embedded in the host language (e.g., COBOL, PL/1) and are accessed usually through CALL statements, but sometimes by extensions in the language.

INDEX - A table used to determine the location of a record. [MART 75]

INTERROGATION - The function of data selection and qualification, extraction, manipulation, and result presentation. [CODA 76]

INVERTED FILE - A file structure which permits fast spontaneous searching for previously unspecified information. Independent lists or indices are maintained in records keys which are accessible according to the values of specific fields. [MART 77]

LOGGING - The recording of actual changes to a data base as updating occurs. [CODA 76]

NETWORK - A set of directed relationships between owner and members such that a single member record may belong to one or more owner relationships.

QUALIFICATION - The specification of selection criteria

during the formulation of a query. [CODA 76]

QUERY - same as interrogation.

QUERY LANGUAGE - Language which permits spontaneous queries to be entered, or allows a non-programmer user to explore the data base and produce reports embodying its data. [MART 75]

PARAMETRIC USERS - Parametric users deal with specific elements of data according to a predetermined procedure and using a limited complement of commands. They are basically unconcerned with data structure and flow except in terms of the relation of data items to identifiers. [CODA 76]

PRIVACY KEY - A facility specified as a literal, a data item, or a procedure used to unlock an operation which is locked. [DATE 75]

PRIVACY LOCK - A facility specified as a literal, a data item, or a procedure used to prevent an operation from proceeding unless the matching privacy key is presented. [DATE 75]

RECORD (logical) - A collection of related values treated as a logical unit during any operation of the database management system (e.g., during data collection, processing, or output).

RECORD (physical) - A unit of data to be placed on, or taken from, a storage device in a single operation.

RECOVERY - The regaining or the bringing back to an original position or condition. [CODA 76]

RELATIONAL - A way of modelling data structures by relations. Relations are usually represented as a collection of tables where each table contains the occurrences of a particular relation. Each column of the relation corresponds to an attribute and each row is an instance of the relation.

SCHEMA (conceptual schema) - A complete description of the database in terms of the characteristics of the data and the implicit and explicit relationship between data-units.

SELF-CONTAINED SYSTEM - A database management system, the capabilities and language of which are intended primarily for the non-programmer. They are self-contained in the sense that they usually have no connection with

any procedural language except that the system itself may be written in a procedural language, or it may permit user-written code in a procedural language.

SUBSCHEMA (or external schema) - A description of those data-units and relationships from a database of interest to a particular program.

UPDATE - The process of changing values in all or selected entries, groups, or data items stored in a data base or adding or deleting data occurrences. [CODA 76]

WORKPLAN

1. Objectives

- o Recommend and justify need for Standards in the Data Description Language area based on Federal government user needs.
- o Develop plan of standards activities to meet such needs.

2. Methodology

- o Define subcommittee scope and objectives
- o Develop definition of DDL
- o Define products of study
- o Define tasks required
- o Develop work plan
- o Implement plan
- o Analyze findings
- o Produce recommendations

3. Tasks

- o Define "Data Description Language"
  - ++Review existing definitions
  - ++Review CODASYL proposal
  - ++Define role of DMCL
  - ++Produce DDL definition
- o Review existing DDL implementations for different data models
  - ++Categorize attribute definitions
  - ++Categorize structure definitions

o Summarize and analyze results of previous tasks

++Review work of other DBMS standards groups

#### 4. Develop conclusions and recommendations

++Review of Data Description Languages. A limited number of DDL definitions were examined. The purpose of the review was to develop a working definition that reflected the scope of the subcommittee area of responsibility.

The definitions reviewed were:

1. Principles of Data Base Management - J. Martin
2. CODASYL - Data Description Language, Journal of Development, June 1973
3. Data Base System, A Practical Reference - Ian Palmer
4. TG-24 Data Base Management System Terminology

#### DDL DEFINITION

A Data Description Language (DDL) is a stand-alone language that defines:

1. Attributes of data elements
2. Logical structure of the data base
3. Logical relationships among units of data (records)
4. Logical methods of access to data

#### THE DDL DEFINITION DOES NOT INCLUDE DMCL

The Device Media Control Language (DMCL) is used to map the data onto physical storage media. It describes the physical location and physical organization of the data. As such, it is the bridge between the DDL and the host DBMS or computer.

#### SEPARATION OF DDL INTO DATA ATTRIBUTES AND DATA STRUCTURES

Attribute Definition:

1. Identifies type of data category such as item, record, file
2. Names items, records, files
3. Specifies sequence of items in records
4. Defines method of data encoding (numeric, binary, ASCII, etc.)
5. Defines length of data items
6. Identifies repeating groups of items
7. Validation or range values for data items

Structure Definition:

- o Identifies key data items
- o Specifies which records are related
- o Specifies how records are related
- o Names data relationships
- o Specifies sequence of records, and files
- o Specifies method that will be used to access data

EVALUATION OF EXISTING IMPLEMENTATIONS

Network, hierarchical and relational DBMS implementations were analyzed for adaptability to DDL standardization. The necessity for different methods of describing data attributes and structure was of special concern.

The CODASYL specification of COBOL type data attributes descriptions was adequate, easy to use, and widely known. Other methods of defining attributes gained nothing for the price of their being different.

Specification of file size or placement of data into physical storage units is dependent on hardware. Each DBMS implementation is designed to its host computer. This area was not considered a candidate for standardization, and removed from our definition of DDL.

Data structures, and associated access facilities, were the most divergent areas. Each structure had its own advantages, depending on an application using the data base. What is good for retrieval may not be good for update; what is good in batch, may not be good on-line.

No single structure was best for all applications. However, the three structures together seem to meet the needs of most applications.

## INTRODUCTION

What is a data base management system? What functions does it provide? These questions must be answered before a standard Data Management Language (DML) model can be developed. After all the functions provided by the set of software identified as data base management systems (DBMS) are qualified, then the question "what can be standardized?" can be answered. The amount of effort required to define DBMS and survey the current market of available DBMS will be considerable. However, it is recommended that data base management be defined concisely and that every function provided by every system that meets the definition be identified. The data management function (DML) subcommittee of Task Group 24 (DATABASE SYSTEMS STANDARDS) has chosen a subset of available DBMS to illustrate how the standards effort could be applied to data base management language functions.

Twelve data management systems were arbitrarily selected for functional analysis. This subset will illustrate optional approaches that are possible with a given set of candidate DBMS. This paper only considers the standardization of host dependent data management languages of the candidate DBMS. The optional approaches leading to a standard DML for data base management systems will be treated. The recommended approach will be emphasized. The following sections of this appendix will cover the approach that was taken to arrive at the recommendations presented in part II of this DML subsection of this report.

## METHOD AND SCOPE

There are many types of data management systems. Most of them can be classified as one of the following types:

1. Data retrieval systems
2. File management systems
3. Complex file systems
4. Teleprocessing monitors
5. Data base management systems

## 6. On-line data management systems

## 7. Special purpose systems

There are subtle and marked differences as well as overlapping capabilities among the types of data management systems. But, each available data system can be categorized into one of these types. TG-24 has been chartered to investigate standardization of data base management systems. The DML subcommittee is limited to evaluating the possible options open to standardization of the data management functions that require the support of a host programming language, such as Fortran, Cobol, etc. It has been found that considering only the DML in standardizing DBMS is not practical. The DML is entwined with the rest of the data base management parts: the DDL, the DMCL, and the host computer operating system. So, the first step the DML subcommittee decided to take was to collect all the data management systems that fit the type "data base management systems." This seems simple enough, except that there are no formal well accepted definitions for any of the above types of data management systems especially for DBMS. TG-24 had two problems. First, there were no acceptable definitions. Second, there are a large number of data management systems available that need to be measured against a definition.

The DML subcommittee of TG-24 felt it necessary to limit the field in order to develop a standard model DML containing a representative set of the data management functions provided by the set of existing DBMS. In order to accomplish this subsetting to one level, a definition of a data base management system in terms of major functions provided was proposed. This definition is not intended to be the final acceptable definition of data base management systems. Its purpose is to limit the available DBMS candidates for the first phase standardization effort. A broader definition that will include all data management systems in the standards effort is recommended after a firm approach is selected using a smaller set of DBMS. Also, for an acceptable definition to evolve, a preliminary one must be provided as a base. The definition for a data base management system in terms of functions and facilities provided is as follows:

1. It is an integrated sharable set of computer programs that perform data access functions.

2. It separates data access (DML) and data definition (DDL) from application programs in a uniform manner.
3. It facilitates and controls in a uniform manner storage and retrieval of data in and from a data base using random access storage devices.
4. It is generalized in the areas of file structure, the number and types of physical files supported, and in number and type of applications supported.
5. It supports and maintains the integrity of the data; validity checking and file backup and recovery utilities.

Now that a definition exists, a survey of the market can begin, and all the data management systems that meet the definition can be selected as candidate DBMS for the DML standardization evaluation. It was found that the number of data management systems that need to be surveyed was too large for TG-24 to handle in the one year time frame given to our initial study of DBMS standardization feasibility. The DML subcommittee chose to select a small subset of systems that are commonly accepted as DBMS to illustrate the options open to DML standardization.

The DML subcommittee chose twelve DBMS for evaluation. The twelve systems are as follows:

1. COMPUTER CORPORATION OF AMERICA - MODEL 204
2. DIGITAL EQUIPMENT CORPORATION - DBMS/10
3. CULLINANE - IDMS
4. HONEYWELL INFORMATION SYSTEMS - IDS/II
5. MRI - SYSTEM 2000
6. SOFTWARE HOUSE, INC. - SYSTEM 1022
7. UNIVAC - DMS/1100
8. IBM - IMS
9. BURROUGHS - DMS/II
10. UNIVERSITY OF CALIFORNIA - INGRES

11. GENERAL MOTORS RESEARCH LABORATORY - RDMS

12. UNIVERSITY OF TORONTO - TORUS/ZETA

Using the twelve DBMS selected, a chart was built showing the major DML commands of each DBMS. See Chart I. Each of the DBMS has other commands for control and interface with respective operating systems that were omitted for this illustration. Examination of Chart I will show that at least four possible categories of command/function sets can be identified by selecting common command subsets. Also, if the data structure models were made known, the categories could be identified by data structure. This is not surprising because the DML commands directly relate to the data structure model of a given DBMS. The four categories that can be identified are network, hierarchical, relational, and table structure. There may be more possible categories. The area of DBMS classification and definition is complex and controversial. To resolve this area satisfactorily may require a separate committee and a year or two's time. A network system is an extension of a hierarchical system. It accommodates hierarchy but also allows complex networks. Most CODASYL DBMS fall into this category. A hierarchical system accommodates data hierarchy in the data structure model, and supplies commands to manipulate hierarchical data. The relational system defines a range of values across a domain as a relation. Relations are defined as tuples involving one attribute and many values. Also, the creation and destruction of relations is dynamic. The self-contained system is one that does not need another language to act as a host. However, many self-contained systems offer host language capability. The difference is in the physical data structure; a self-contained system usually has a static physical data structure, such as a table structure. Regardless of the logical structure of the data base, the physical structure is the same in most self-contained systems.

DBMS designs have different purposes; i.e., relational systems allow dynamic changes to be made to sets and relationships to handle a moderate amount of volatile data. The CODASYL approach allows very flexible data structure for large amounts of data, but the structure is very static once defined. The command sets that accompany the different types of data base management systems are directly related to the data models of the particular type. This is the reasoning behind categories of data base management systems that are illustrated in Charts A, B, C, and D.

Both the common and unique commands are illustrated in Charts A, B, C, and D. The DBMS were grouped by evaluating common commands, unique commands, and data models. The names given to the resulting categories are not important, and this is another area that could be standardized. Some systems could fit into more than one category, such as System 2000. System 2000 is a table structure system, but it is also a hierarchical system. It was included in both Charts B and D to illustrate dual category possibilities. This also provides encouragement for the possibilities of DBMS evolving to provide similar command/function capabilities. Network and hierarchical systems are similar enough to evolve to one category in the near future. There is probably one other category that will result if a more complete comparison and classification effort is undertaken. It will be the special purpose data base management system category. In some cases, semi-generalized data base management systems are developed for special purpose hardware and/or applications. Due to the special purpose nature of such a DBMS, it would be very difficult to standardize. One special purpose system is usually not very similar to another. So, this category would hold a group of DBMs that are dissimilar to themselves and the other categories and not amenable to standardization.

#### PROCESS LEADING TO DML RECOMMENDATIONS

Now that the DBMS are categorized into subsets, some options for standards can be proposed. Three or four options are possible today. First, a standard DML interface language for each category could be developed. Second, an interface language could be developed to include all identified categories; one standard interface language with preprocessors to generate individual DML's. Third, a set of standard DML commands could be developed for each category identified from the analysis of all candidate DBMS. Of course, the option of no recommended standards at this time is a possibility. The possibility of selecting an existing DBMS from a standard appears unlikely because no single DBMS will have a super set of commands/functions that would be needed to provide all the capability now available by use of a combination of available DBMS.

##### 1. Multiple Interface Option

Most competitive DBMS will have the same major capabilities within the next five years. Of course, the syntax and semantics of the interface languages (host language interfaces) will not be standard unless reasonable standards are proposed and accepted. The standard interface approach for categories of DBMS provides a flexible method for gaining applications program DBM standardization. When, and as DBMS

begins to provide the same functions and capabilities, the interface(s) can be modified to accommodate changes or additions to the standard DMLS. Eventually, a single standard DML interface may evolve if the DBMS becomes similar enough. Vendors of DBMS are already providing similar capabilities to remain competitive. The users should begin to be concerned with standard methods of identifying capabilities before the vendors implement new capabilities.

## 2. Single Interface Option

The option of beginning with a single standard DML interface may be possible, but it would be very complex and may not be acceptable to the end user. It would be necessary for the single interface to accommodate the full set of DML commands of all the candidate DBMS selected. This would result in command translations that would cause no operations in any DBMS that do not provide the command capability desired. It is recommended that if the interface option to standardization is selected, the first option chosen should be one standard interface for each identified category. If this approach is followed, then, as the DBMS categories provide the same command/functions, one standard interface may evolve. One advantage to interface approach is that the user can implement or have the interface implemented without cooperation from the vendor of any or all the candidate DBMS. However, vendor cooperation is desired if possible. It is also recommended that the feasibility of developing a single DML standard for all host programming languages for a given category of DBMS be evaluated.

## 3. Standard DML Language Option

A set of DML commands could be specified for each category of DBMS. Again, it is wise to stay within the categories to avoid translations that would result in no operations. Chart I gives an indication of the difficulties that would be faced in developing a standard DML command set for all available DBMS. This approach would not provide the user with a working standard as readily as the interface approach, because after the standard language is specified, the individual vendors of each DBMS in each category would be responsible for implementing the standard DML, if indeed they would be agreeable to do so. The standard DML would not insulate the user from changes in individual DBMS as well as in the interface option. Changes to the DML of a given DBMS could be handled in the standard interface instead of in each application program. If changes were made to a standard DML system, each application that would be affected would require change. However, this option would allow an application program to access any DBMS of a given category in a computer network using one DML. For a given

category, see Charts A, B, C, and D, the major DML commands are already similar. The problem with specifying a standard DML, is that each vendor of each candidate DBMS for each category must implement the standard. This is a big problem because of cost to the vendor and the fact that one or more of his major capabilities may not be standard. This could make vendors unwilling to agree to standardize their DML. Anyway, it is recommended that a standard DML set be specified for each category of DBMS for two reasons. One, by specifying the DML set, the data management functions for each category are identified. Two, the DML sets for each category are basic to whatever approach is taken to develop a DML standard.

#### 4. No Standard DML

No standard is always a possibility, but in the case of data base management systems a lack of standards is costly and is going to be more costly. Hardware advances in the next decade will probably provide more speed to conventional DBMS rather than new DBMS concepts. Standardization of current conventional DBMS functions will not only help the users within the next decade, but will help the designers of future data base management systems by emphasizing the major DML functions through standardization. Then only new system capabilities will need standards attention.

#### BENEFITS EXPECTED

It would be very difficult to place a dollar value on DBMS standardization versus no standardization. It is a field of computer science that is still in infant stages. Most installations have only recently implemented their first production system using a DBMS. Consequently, there is little data available for measuring conversion or training costs in regards to standard or non-standard DBMS. It can only be speculated to be expensive to convert applications from one DBMS to another. Also, it can be speculated that some conversions may not be necessary if the same application program could access data of two or more DBMS without modifications. This implies benefits for standard DBMS that cannot be proven as yet.

An important benefit to be gained from the standardization of DBMS will be related to the heterogeneous computer network environment and distributed data bases. Currently, the hardware and software exists to provide the communications capability for a heterogenous computer network. However, data bases under the control of different DBMS on the various computer nodes of the network cannot be readily accessed in a uniform or standard manner. A standard DML interface or a standard DML would reduce the programming and

training effort necessary to access data controlled by different DBMS. It would be possible for a single application program to access similar data under the control of different DBMS without the need to change or rewrite the program. Of course, other system software, such as operating system interfaces, user protocols, etc. would also require standardization to allow the DML standard to be fully and effectively used. Changes to the system software could be handled in the interfaces and free applications programmers of any concern. Maintenance could be centrally handled by updating and changing the interfaces a large percentage of the time instead of updating each individual application program. Life cycle support of applications systems could be greatly reduced.

## DEVELOPMENT AND IMPLEMENTATION FEASIBILITY

Development possibilities definitely exist for the standard interface approach. The best thing the approach has going for it is the fact that one vendor or large user could design and develop the interfaces. Most other approaches rely on the individual vendors of each DBMS to implement a proposed standard. This is fine, but an automatic method of sharing data between heterogenous systems is not obtainable through a multi-vendor standards effort. For instance, the CODASYL standard provides a standard specification for a data base management system; however, two different vendor implementations are not compatible for data sharing purposes. The languages are compatible: the DDL and the DML. An application program compiled with the DML of one CODASYL system cannot access the data belonging to another CODASYL system without being rewritten and recompiled. If the application program were to use a standard DDL and DML interface, the interface could handle the translation to a different data management system. As previously stated, an interface could be developed for each category of DBMS identified. This is recommended as a first phase DBMS standardization effort.

The proposed interface would consist of a precompiler stage plus an execution time management routine. The precompiler would generate a parametric call to a generic data management routine. At execution time, depending upon which DBMS is to be invoked, a specific linkage could be made. The DDL methodology would also require a standard interface to handle dynamic linkage to specific data structures. This is anticipated to be a later phase of the interface standardization effort. The first level will be to standardize the DDL and DML interfaces for each identified category of DBMS. An application program written using one of the standard sets of interfaces would be precompiled generating calls to a specific DBMS. Once an application

program is compiled, it would be locked into a specific data base management system. However, it would only need to be recompiled using a different precompiler to access a different DBMS of the same category. This approach would provide the applications user with a standard DML set for each category of DBMS in the near future, and very possibly evolve to one set if the DBMS becomes similar enough in the short range.

## MAINTENANCE

The standard interface approach insulates the user from vendor changes to the given category of DBMS he is concerned with. If a data management command/function does not exist in a particular DBMS, the interface could emulate the function or notify the user that the feature is not currently available through the given DBMS he is trying to access. When, and if, the function becomes available, the above procedure could be dropped, or if the function was never available in the interface, it could be added. The user would not be required to modify existing DBMS host applications programs every time a change occurred to a given DBMS.

The maintenance functions would be transferred from the individual applications programmers to a central staff of data base system administrators (DBSA). The time and personnel costs could possibly be reduced depending upon the number of application areas supported by a given data base management system. With a large number of unique applications supported by a given DBMS, a saving could be realized by centralizing the support rather than dispersing the maintenance activities over all the application areas to be supported. In the case where only one special purpose application is to be supported, the savings would not occur.

## CONCLUSIONS

Standardization of the data management function is definitely recommended. Many approaches are possible, but the interface is probably the most viable one. It does not rely on the efforts of each candidate DBMS vendor to be implemented and it guarantees not only syntactic compatibility but compile and execution time compatibility for at least a given category of DBMS. It is not very likely that individual vendors of various DBMS will ever reach the point where they will offer systems that could provide the same level of compatibility that a standard set of interfaces could.

Charts A, B, C, and D illustrate the elementary command/function sets that would form the basis for the standard interface languages for the four categories identified in this paper. There may be more categories identified

when a complete survey is done using a broader definition for a DBMS. There will definitely be more commands and functions identified for inclusion in the standard interface languages for each category of DBMS that is identified. This approach is not limited to commercially available DBMS. An in-house system could be included by building an interface to it. Another sometimes important advantage to the interface approach is that the existing data bases and applications can remain unchanged while new programs and data bases are designed and written.

The goals of this proposed standardization effort should be to provide the user with a standard and uniform method to build and access data bases using any one of the data base management systems from a given category. It would be ideal to be able to provide only one access or interface language for all data base management systems. However, that is not practical today. The team or committee that does the survey of data base management systems would be wise to investigate the facilities and functions needed by the generalized programming user and specify the characteristics of such an interface and measure the existing systems against it. Also, in an effort to develop one single interface language for DBMS, it would be wise to investigate the correspondence between the relational calculus function of relational DBMS and the steps of the conventional inquiry process (i.e., self-contained, CODASYL, hierarchical).

DMF comparison charts I, A, B, C, & D

CHART I DBMS SYSTEMS  
BASE LINE DML COMPARISON

SYSTEMS

Commands	M204	S1022	DBMS10	DMS1100	IDMS	IDSII	DMSII	S2F	INGRES	TORUS/ZETA	RDMS	IMS
OPEN	y	y	y	y	y	y	y	y	n	n	n	n
CLOSE	y	y	y	FINISH	FINISH	FIN	y	y	n	n	n	n
FIND	y	y	y	y	y	y	n	y	RETRIEVE	GFT	GET	n
GET	y	y	y	y	v	n	y	y	n	y	y	n
FETCH	n	n	n	y	y	y	y	n	n	n	n	GU/GN
IF	n	n	y	y	y	y	y	n	n	n	n	n
STORE	y	ADD	y	y	v	y	y	ADD	APPEND	PUT	PUT	ISRT
INSERT	n	n	y	CONNECT	CON	CON	y	y	n	y	n	ISRI
REMOVE	n	n	y	DISCON	DISC	DISC	y	y	n	n	n	DLET
MODIFY	CHNG	CHNG	y	y	y	y	y	y	REPL	UPDATE	MODIFY	REPL
DELETE	y	y	y	ERASE	ERASE	ERASE	y	y	y	y	y	DLET
SET	n	n	y	y	n	y	y	n	n	n	n	n
INVOKE	y	y	y	y	RIND	y	y	y	n	n	n	n
ACCEPT	n	n	MOVE	y	y	y	n	n	n	n	n	n
CREATE	n	n	n	n	n	n	n	n	y	y	y	PUT
DESTROY	n	n	n	n	n	n	n	n	y	y	y	n

SYSTEMS

Commands	DBMS/IM	DMS/1100	IDMS	IDS/II	Function
READY	OPEN	READY	READY	READY	Open a data base file
FINISH	CLOSE	FINISH	FINISH	FINISH	Close a data base file
FIND	FIND	FIND	FIND	FIND	Indexed search
Calc	yes	yes	yes	yes	Hashed find
Direct	yes	yes	yes	yes	Direct key find
Owner	yes	yes	yes	yes	Owner of set find
Record	yes	yes	yes	yes	Specific record name find
GET	GET	GET	GET	GET	Get record values
FETCH	no	FETCH	OBTAIN	no	Find and get combined
STORE	STORE	STORE	STORE	STORE	Place a record occurrence in DB
CONNECT	INSERT	CONNECT	CONNECT	CONNECT	Add a record to a specific set occurrence
DISCONNECT	REMOVE	DISCONNECT	DISCONNECT	DISCONNECT	Remove a record occurrence
MODIFY	MODIFY	MODIFY	MODIFY	MODIFY	Change field values of selected records
ERASE	DELETE	ERASE	ERASE	ERASE	Delete record occurrences
ACCESS	ACCESS	no	no	no	Access a given subschema
INVOKE	INVOKE	INVOKE	HIND	INVOKE	Link program to given schema/subschema
ACCEPT	MOVE	ACCEPT	ACCEPT	ACCEPT	Read currency status indicators
USE	USE	USE	USE	USE	Specify error procedure/routine
IMPART	no	yes	no	no	Link to run time DBMS control program
DEPART	no	yes	no	no	Break link to DBMS
EMPTY	yes	no	no	no	Fortran function to test for empty set
MEMBER	yes	no	no	no	Fortran function to test for member
OWNER	yes	no	no	no	Fortran function to test for owner
TENANT	yes	no	no	no	Fortran function to test for tenant
COPY	no	no	yes	no	Include DBMS procedures in host program

CHART B Hierarchical Data Mode)

SYSTEMS

Commands	IMS	DMS/II	SYSTEM 2000	Function
.....	.....	.....	.....	.....
INVOKE	no	INVOKE	SCHEMA IS	Attach to specific data base
.....	.....	.....	.....	.....
ASSIGN	no	ASSIGN	ASSIGN	Links bet records or values to existing trees
.....	.....	.....	.....	.....
OPEN	no	OPEN	OPEN	Open specific data sets
.....	.....	.....	.....	.....
CLOSE	no	CLOSE	CLOSE	Close specific data sets
.....	.....	.....	.....	.....
DELETE	DLET	DELETE	DELETE	Delete record occurrences
.....	.....	.....	.....	.....
REMOVE	DLET	REMOVE	REMOVE TREE	Delete trees from data base
.....	.....	.....	.....	.....
FIND/GET	GU/GN	FIND/GET	GET/LOCATE	Sequential or indexed sequential find
.....	.....	.....	.....	.....
FIND KEY	GHN/GNP	FIND KEY	GEID/GETA	Indexed find of records, trees, or segments
.....	.....	.....	.....	.....
GENERATE	no	GENERATE	no	Set creation at execution time
.....	.....	.....	.....	.....
INSERT	ISRT	INSERT	INSERT	Inset a tree in data base
.....	.....	.....	.....	.....
STORE	ISRI	STORE	ADD	Store individual records in DA
.....	.....	.....	.....	.....
SET	no	SET	no	Alter record or set currency
.....	.....	.....	.....	.....
CREATE	PUT	CREATE	LOAD	Load data into a defined DH structure
.....	.....	.....	.....	.....
MODIFY	REPL	MODIFY	MODIFY	Change fields of record(s)
.....	.....	.....	.....	.....
DESCRIBE	no	DESCRIBE	DESCRIBE	Printout of data structure
.....	.....	.....	.....	.....

CHART C Relational Model  
SYSTEMS

Commands	INGRES	TORUS/ZETA	RDMS	Function
RETRIEVE	RETRIEVE	GET	GET	Find relation <cond> <access method>
APPEND	APPEND	PUT	PUT	ADD tuples to a relation
REPLACE	REPLACE	UPDATE	MODIFY	Change tuple values
CREATE	CREATE	CREATE	CREATE	Create a relation for a range or domain
DESTROY	DESTROY	DESTROY	DESTROY	Destroy a relation
DISPLAY	no	no	DISPLAY	Printout of relation structure
INDEX	INDEX	?	?	Create a secondary index for a relation
RANGE	RANGE	NO	NO	The range of values for a relation
MODIFY	MODIFY	MODIFY	no	Change access method for a relation
SUBSET	no	no	SUBSET	Subset a relation
COMBINE	no	no	COMBINE	Combine relations
CREATEDB	CREATEDB	CREATEDB	CREATEDB	Create a data base
DESTROYDB	DESTROYDB	DESTROYDB	DESTROYDB	Destroy a data base
DELETE	DELETE	DELETE	DELETE	Remove values of a relation

CHART D Table Structure Model

SYSTEMS

Commands	M204	SYSZK	SYS1022	Function
.....	.....	.....	.....	.....
OPEN	IFOPEN	yes	DROPEN	Open a data base file
CLOSE	IFCLOSE	yes	DPCLOSE	Close a data base file
FIND	IFFIND	WHERE	DAFIND	Indexed find
LOAD	IFFLOD	LOAD	DLOAD	Load data base
SEARCH	no	yes	DASRCH	Sequential retrieval
SAVE	IFFILE	SAVE	DFAVE	Save record pointers of found set
COUNT	IFCOUNT	COUNT	DBNREC	Count records found
SORT	no	yes	DASORT	Sort records found
DELETE	IFDREC	DELETE	DBDEL	Delete records found
CHANGE	IFPUT	CHANGE	DACHNG	Change field values of records found
ADD	IFSTORE	ADD	DBADD	Add new records
APPEND	IFFLOD	APPEND	DRADDR	Add a group of new records
ALLOCATE	no	no	DRALLO	Increase disk allocation of DR
KEY/UNKEY	no	yes	DRKEY	Create or recreate new key table for field
INFORM	no	DESCRIBE	DRINFO	Display data structure for DR
OPTIMIZE	no	yes	DROPT	Optimize key tables
MORE	IFMORE	yes	no	Continue from previous retrieval procedure
CHKPT	IFCHKPT	yes	no	Enable checkpoints
DIAL	IFDIAL	no	no	Set up link to on-line system
ENO/DEQ	IFENO	yes	no	Initiate enqueue/dequeue facility
GET	IFGET	yes	DARGET	Get values for records found
CALL	IFCALL	no	DBCALL	Call other system programs and return
FINISH	IFFNSH	EMDS2K	DREND	End link to DRMS
START	IFSTRT	STS2K	DFFIO	Initiate link to DRMS

## 7.6 APPENDIX 6 - DATA DICTIONARY/DIRECTORY

### ++Definition.

A data element dictionary/directory (to be referred to in short as data dictionary (DD)) is a software tool that is used to control the totality of data elements within an application. It is viewed as the central repository of all descriptive information about each data element contained within an application data base. It lists, describes, and locates each data element in a data base. The dictionary does not manage the actual content of the data, but it does manage the descriptive characteristics of data.

The dictionary portion is a glossary of terms of data elements representing their characteristics and logical relationships with data base components and application usage. The dictionary describes what data are contained in the organization's data base.

The directory portion contains object data definitions as used by the computer, plus physical storage locations and access strategies. The directory locates where data are stored.

### ++Usage.

The DD serves the database administrator, systems analyst, software designer, and programmer by providing a central repository for information about data resources. It aids people in planning, controlling, and evaluating the collection, storage, and use of the data resources.

As a documentation tool, it captures descriptive information and aids in establishing standards for data naming, usage, and coding conventions. It identifies users or programs which may be affected by any changes, additions, or deletions to the data.

If the DD is interfaced into the structure of a DBMS, then it may be used for generating the data description language of the DBMS, maintaining data descriptions and compiling or interpreting program references which use these descriptions.

## ++Functions of Data Dictionary.

The primary function of a DD is to provide a method of centralized control over data elements. The following are some functions of existing DD. No one system performs them all.

- \* Data attribute - Name, Length, Type
- \* Data relationship - The structural properties among data
- \* Synonyms - Allow alias capability
- \* Textual description
- \* Access control specification
- \* Edit and validation check rules - e.g., null and default values, ranges of value permitted, etc.
- \* Output decoding specification - coded values to be translated so that it is human-readable.
- \* Key/Non-key - Indicate whether the data element is meant to be searchable or not.
- \* Physical storage specification - logical and physical address, internal representation, compaction, etc.
- \* Usage information - cross-references to users, application programs, and output reports.
- \* Frequency of access - usually used for statistical monitoring purposes.

## ++Types and Commercial Dd's.

There are at present at least four broad groups of data dictionary systems existing in the market that can be identified:

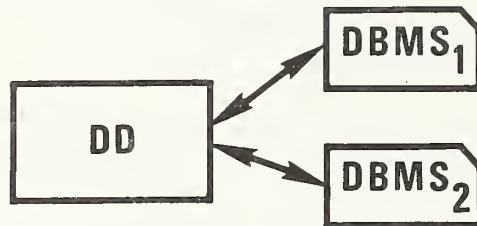
\* Type A - The free-standing package which could be used in a non-DBMS environment.



Examples of commercial DD are:

DICTIONARY	SUPPLIERS
DATA CATALOGUE DATAMANAGER LEXICON	Synergetics Corp. Management Systems & Programming, Ltd. Arthur Anderson & Co,

\* Type B - The free-standing package that optionally provides interfaces to one or more DBMS.



Examples of commercial DD are:

DICTIONARY	SUPPLIERS	DBMS INTERFACE
DATA CATALOGUE	Synergetics Corp	IMS, TOTAL
DATAMANAGER	Management Systems and Programming, Ltd.	ADABAS IDMS, IMS, TOTAL
LEXICON	Arthur Anderson & Co.	IDMS, IMS, TOTAL

\* Type C - The single data dictionary package designed to co-exist with a particular DBMS. This type of data dictionary is solely dependent on the DBMS and usually developed and marketed by the same vendor.



Examples of commercial DD packages are:

DICTIONARY	SUPPLIER	DBMS REQUIRED
Control 2000	MRI Systems	System 2000
DB/DC Dictionary	IBM	IMS or DL/1
UCC TEN	University Computing	IMS
Data Control System	Haverly Systems	DMS-1100
Data Dictionary	Cincom Systems	TOTAL
IDD	Cullinane	IDMS

\* Type D - The data dictionary that is embedded within a DBMS. The data dictionary function usually is part of the data definition function and the meta-data is stored as part of the database for the DBMS.



Examples of commercial DD are:

DICTIONARY	SUPPLIER	DBMS REQUIRED
GIM II Dictionary	TRW	GIM II

## ++Information In a Data Dictionary System.

A data dictionary captures information about data elements. Some of this information seems to be provided by all data dictionary packages and deemed essential while others, though desirable, may be omitted at the discretion of the implementor.

To concentrate on the information requirements of those involved in data and database administration, several types of information usage were identified. Certain information is captured to be used by people for documentation and control purposes, while other information is necessary for the generation of data definition language, data manipulation language and application program processors. For illustrative purposes the following table shows the information captured in a data dictionary and its uses by various users.

	People	DDL	DML	Application Processor
Information in DD				
Name, Type, Length		X		
Structure Relationship		X	X	X
Synonyms	X			X
Textural Desc.	X			
Access Control		X		X
Edit & Validation				X
Key/Non-Key		X		X
Usage Info.	X			X
Freq. of Access	X			
Physical Storage Specification				X

## 7.7 APPENDIX 7 - END-USER/QUERY LANGUAGE

The following subjects were considered when preparing the End User Facility recommendations.

### ++List of End User Facilities.

1. Report Specification Language
2. Enquiry Specification Language
3. Update Specification Language
4. Parametric Interface

### ++List of Classes Of Users.

1. System Analysts/DBA's Staff
2. Application Programmers
3. On-line Job-trained User
4. Researcher
5. Casual User

### ++Measures of Query Languages.

1. Quantitative Measures
  - a. Level
  - b. Completeness
2. Qualitative Measures
  - a. Mathematical Sophistication
  - b. Learnability
  - c. Procedurality

### ++List of Variable Features In End User Facilities.

- A. Procedurality
  1. Procedural
    - a. logic
    - b. navigation
  2. Non-procedural
    - a. form fill
    - b. question/answer
    - c. menu
- B. Report formatting
  - a. flexible
  - b. pre-formatted
  - c. automated display of draft format

- d. construction assistance
- C. Request Analysis
  - a. monolithic
  - b. incremental
- D. Level
  - a. % code reduction vs. COBOL (keystrokes)
  - b. % coding time reduction vs. COBOL
- E. Default treatment
- F. Ratio of simplicity to functionality
- G. Retrieval criteria permitted
  - a. mathematical operators
  - b. logical operators
  - c. string operators (begins, contains)
  - d. set functions (count, sum, mean, standard deviation, maximum, minimum)
- H. Cascading retrieval
- I. Extensibility
- J. Processing of conditionals
- K. Update capability
  - a. within queries
  - b. similar to query syntax
  - c. special syntax
  - d. transaction oriented
  - e. form-fill
- L. Navigation
  - a. DBA intervention required to change access paths
  - b. fixed path chosen
  - c. multi-path ambiguity negotiated
- M. Execution efficiency
- N. Aggregation operations
  - a. sort flexibility
  - b. sort efficiency
  - c. data clumping capability
- O. Reusability of user constructs
  - a. simplicity of retention
  - b. simplicity of temporary modification
  - c. simplicity of permanent modification
- P. Data type support

- a. different types handled
  - b. conversion support
  - c. ease of addition of special types
- Q. Fog index
- a. documentation
  - b. syntax
- R. Benchmarking

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBS SP-500-51	2. Gov't. Accession No.	3. Recipient's Accession No.
4. TITLE AND SUBTITLE RECOMMENDATIONS FOR DATABASE MANAGEMENT SYSTEM STANDARDS		5. Publication Date August 1979	6. Performing Organization Code
7. AUTHOR(S) FIPS Task Group on Database Management Systems Standards, J. Berg, Chairman		8. Performing Organ. Report No.	
9. PERFORMING ORGANIZATION NAME AND ADDRESS NATIONAL BUREAU OF STANDARDS DEPARTMENT OF COMMERCE WASHINGTON, DC 20234		10. Project/Task/Work Unit No.	11. Contract/Grant No.
12. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (Street, City, State, ZIP)		13. Type of Report & Period Covered Final	14. Sponsoring Agency Code
15. SUPPLEMENTARY NOTES Library of Congress Catalog Number: 79-600087 <input type="checkbox"/> Document describes a computer program; SF-185, FIPS Software Summary, is attached.			
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.) <p>In March 1977, FIPS Task Group 24 initiated a study of the need for database standards within the Federal government. The voluntary participants from several Federal agencies considered the actions of other standards bodies; reviewed the alternatives to Federal standards; examined the issues of standards adoption, timing, and impact on technology; developed a method for justifying standards, and attempted to anticipate likely database technology advancements.</p> <p>TG-24 recommended standards in certain specific technical areas, concluded that standards were premature in others, and emphasized the need for certain guidelines.</p> <p>This final report of TG-24 contains the recommendations for standards and guidelines as well as the assumptions, benefits, and costs considerations used to justify the recommendations.</p>			
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) Database; DBMS; data-description; data-dictionary; data-directory; data-manipulation; languages; query; standards.			
18. AVAILABILITY <input checked="" type="checkbox"/> Unlimited  <input type="checkbox"/> For Official Distribution. Do Not Release to NTIS  <input checked="" type="checkbox"/> Order From Sup. of Doc., U.S. Government Printing Office, Washington, DC 20402, SD Stock No. SN003-003-02095-8  <input type="checkbox"/> Order From National Technical Information Service (NTIS), Springfield, VA, 22161	19. SECURITY CLASS (THIS REPORT) UNCLASSIFIED	21. NO. OF PRINTED PAGES 99	
	20. SECURITY CLASS (THIS PAGE) UNCLASSIFIED	22. Price \$3.75	



# NBS TECHNICAL PUBLICATIONS

## PERIODICALS

**JOURNAL OF RESEARCH**—The Journal of Research of the National Bureau of Standards reports NBS research and development in those disciplines of the physical and engineering sciences in which the Bureau is active. These include physics, chemistry, engineering, mathematics, and computer sciences. Papers cover a broad range of subjects, with major emphasis on measurement methodology, and the basic technology underlying standardization. Also included from time to time are survey articles on topics closely related to the Bureau's technical and scientific programs. As a special service to subscribers each issue contains complete citations to all recent NBS publications in NBS and non-NBS media. Issued six times a year. Annual subscription: domestic \$17.00; foreign \$21.25. Single copy, \$3.00 domestic; \$3.75 foreign.

Note: The Journal was formerly published in two sections: Section A "Physics and Chemistry" and Section B "Mathematical Sciences."

### DIMENSIONS/NBS

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, \$11.00; Foreign \$13.75

## 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 PUB)**—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, \$25.00; Foreign, \$30.00.

**Liquefied Natural Gas.** A literature survey issued quarterly. Annual subscription: \$20.00.

**Superconducting Devices and Materials.** A literature survey issued quarterly. Annual subscription: \$30.00. 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

---