JPL PUBLICATION 81-88

Survey of Standards Applicable to a Database Management System

Jose L. Ureña

(NASA-CR-165009) SURVEY OF STANDARDS APPLICABLE TO A DATABASE MANAGEMENT SYSTEM (Jet Propulsion Lab.) 51 p HC A04/MF A01 CSCL 09B

N82-13766

Unclas
G3/61 08456

October 15, 1981

NASA
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

JPL PUBLICATION 81-88

Survey of Standards Applicable to a Database Management System

Jose L. Ureña

October 15, 1981

NASA
National Aeronautics and
Space Administration
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

ABSTRACT

The objective of this document is to identify existing industry, government and NASA standards, and the status of some standardization activities of standards-setting organizations, that can be applicable to the design, implementation and operation of a database management system (DBMS) for space-related applications. This document also contains, when applicable, an assessment as to the applicability of the standards that have been identified to a general-purpose, multimission database management system to be called the Applications Database Management System (ADBMS) for use in future missions of the Office of Space and Terrestrial Applications (OSTA) of NASA.

PRECEDING PAGE BLANK NOT FILMED

TABLE OF CONTENTS

SECTION 1

1. STANDARDS

1.1 Introduction

1.1.1 Objective

The objective of this document is to identify existing standards and the status of some standardization activities of standards-setting organizations that can be applicable to the design, implementation and operation of a database management system for space-related applications.

This document contains a compilation of standards with a description of the nature and status, as of August 1981, of some relevant on-going standardization efforts. It also contains, when applicable, an assessment as to the applicability of the standards that have been identified to a general-purpose, multimission database management system to be called the Applications Database Management System (ADBMS) for use in future missions of the Office of Space and Terrestrial Applications (OSTA) of NASA.

1.1.2 Organization of This Document

This document is organized by areas for which applicable standards may exist. For each area, any and all standards or standardization activities that can be relevant and have been identified for that area are described under the same heading.

1.2 Standards Organizations

Within the United States, standards are developed and published on a voluntary basis, since there exists no governmental agency with direct control over the use of standards within the computer (or any other) industry. Further, there exists no congressional authorizations or appropriations that directly support or fund the development of standards within the United States. However, the National Bureau of Standards (NBS) (a division of the Department of Commerce) has direct responsibility for the development of standards for use within the federal government and for compliance with those standards by vendors of equipment to the federal government. Since the majority of computer equipment manufacturers are vendors not only to the United States Government, but also to all consumers and users (nationally and internationally), federally enforced standards can be expected to be *de facto* industry standards, merely to save the cost of producing two separate lines of equipment. While the NBS has the authority to develop independent federal standards, the recognition of what would be the overall costs of producing special custom-designed equipment for the government requires the Bureau to actively participate in and promote the voluntary standards effort of the computer industry.

1.2.1 Federal Government

The following types of practices and standards have been identified in the federal government for data elements and representations:

1) De facto Practices. Those data elements and representations in current use that have not been subjected to official or formal standardization.

2) Unit Standards. Those data elements and representations that have been approved by an authorized official for use within that unit. (A unit is any federal organization within the executive branch of the government, which is at a lower organizational level than an executive department or independent agency.)

3) Agency Standards. Those data elements and representations that have been approved by an authorized official for use within an executive department or independent agency.

4) Federal Program Standards. Those data elements and representations that have been approved by the Secretary of Commerce for use in a particular program or mission where more than one executive branch department or independent agency is involved with their use. For example, those standards that could be approved and prescribed for use are those which include, but are not limited to, federal-wide personnel, communications, and transportation data systems.

5) Federal General Standards. Those representations that have been approved by the Secretary of Commerce for federal-wide use by executive departments and independent agencies in all federal-wide programs and for use in all federal data systems. For example, this includes such representations as calendar dates, state abbreviations and codes, and codes for standard metropolitan statistical areas.

Federal general standards are the highest level standards, followed by federal program standards, agency standards, and unit standards, in that order. This order establishes a precedence for standards use.

Standardization activities within the federal government are the following:

Interagency Committee on Automatic Data Processing (IAC/ADP)

The IAC/ADPs membership consists of 48 federal agencies, and it provides a medium for the exchange of information on management and technological developments, serves as a forum for the discussion of policies and regulations being proposed by the central management agencies, and initiates studies on matters on which the Committee wishes to formulate views and recommendations for consideration.

Federal Information Processing Standards (FIPS) Task Groups

The FIPS Task Groups are composed of technical personnel with a knowledge of each agency's requirements, who assist the NBS in matters relating to the development, adoption, and implementation of standards and in providing better coordination of the Federal Automatic Data Processing (ADP) Standards Program.

Federal Programs Standards Groups for Data Elements and Codes:

1) Federal Telecommunications Program Standards Committee. The National Communications System (NCS) is designated as the lead agency to be responsible for the development and maintenance of federal program standards regarding data elements and codes in Federal Telecommunications Systems.

2) Federal Telecommunications Standards Committee (FTSC). The objectives of the FTSC are: (1) to develop and coordinate standards required to achieve operational compatibility among networks of the National Communications System; (2) in concert with the National Bureau of Standards, to develop and coordinate standards for data transmission and the computer-telecommunication interface; and (3) to increase the cohesiveness and effectiveness of the federal telecommunication communities influence on national/international standards program and on the FIPS program.

1.2.2 National Organizations

The American National Standards Institute (ANSI) is the national clearinghouse and coordinating agency for voluntary standards in the United States [1]. It is a nonprofit federation of approximately 140 trade associations and professional societies, and over 750 companies, which are dues-paying members.

As the national clearinghouse for standards, ANSI provides the machinery for developing and approving standards that are supported by a national consensus. Its constitution states: "In standardization practice a consensus is achieved when substantial agreement is reached by concerned interests according to the judgement of a duly appointed authority. Consensus implies much more than the concept of simple majority, but not necessarily unanimity".

ANSI is the United States member body of the International Standards Organization (ISO) (see Section 1.2.4).

The federal government is a major contributor to the work at ANSI. The Director of the National Bureau of Standards is a member of its Board of Directors. Representatives from various government departments and agencies participate through the many ANSI councils, boards, committees, and task groups.

The Information Systems Technical Advisory Board (ISTAB) of ANSI is responsible for all aspects of standardization of systems that transmit, store, or process information, and advises, among others, the American National Standards Committee X3 (ANSC-X3) for Computers and Information Processing.

ANSC-X3 has the task of standardization related to systems, computers, equipment, devices, and media for information processing systems. The Committee is sponsored by the Computer and Business Equipment Manufacturers Association (CBEMA). As the sponsor, CBEMA acts as the secretariat providing administrative support and, through its Standards Committee, is responsible to ANSI for the general administration of ANSC-X3.

ANSC-X3 accomplishes its responsibilities through two major committees: Standards Planning and Requirements Committee (SPARC) and Standards Steering Committee (SSC).

SPARC is the research and study arm of ANSC-X3, responsible and responsive to the ANSC-X3 committee for the identification of the needs and requirements of the industry for standards. Having identified the need, justified the work, confirmed the availability of resources, and determined that the work is within the limits of current technology, SPARC recommends to ANSC-X3 the establishment of a technical committee under the supervision of SSC for the actual standard development. Once approved by ANSC-X3, the SSC establishes the technical committee and oversees its work.

Although ANSI is responsible for the coordination of national voluntary standards in the United States, the Institute has never established itself as the sole organization for the development of standards-related information. In fact, over one-third of the standards published by ANSI originated from outside the ANSI organization. The list of approved American National Standards developed by ANSC-X3 is given in [2]. Reference [3] provides the current, approved organization membership of ANSC-X3.

1.2.3 Conference on Data Systems Languages (CODASYL)

CODASYL is not a standards-setting body, but instead an informal and voluntary organization specifically established to design and develop techniques and languages to assist in data systems analysis and implementation. Specifically, CODASYL operates four committees: the Data Definition Language Committee, the Programming Language Committee, and the Systems Committee. The overall organization is supervised by the Executive Committee, and there are a number of task groups working on specific issues.

The most widely known products of CODASYL are: (1) the COBOL language, which was subsequently accepted as the basis for the American National Standard Cobol X3.23; and (2) the Database Reports, which have been accepted by ANSI, and which are the basis for a standard that will be published very soon (see Section 2).

1.2.4 International Organizations

The International Organization for Standardization (ISO) was established in 1947 to promote the development of standards, and its objectives, as specified in its constitution, are: ". . .to facilitate the coordination and unification of the standards of Member Bodies." In connection with this goal, ISO may ". . .set up international standards provided that in each case no Member Body dissents."

Present membership in ISO includes 54 member bodies. A member body is an organization of an individual nation which best represents the standardization activities of its nation. The ISO Member Body that represents the United States is the American National Standards Institute (ANSI). The United States' viewpoints, to be presented in the technical work of ISO, may be developed through the interested ANSI committee, through a competent committee of another standards organization, or through a committee specifically organized as an advisory committee to an ISO Technical Committee. The work of the technical committees eventually results in ISO Draft International Standards (DIS), which may be embodied in the national standards of ISO member bodies. Conversely, national standards of the member bodies may be embodied in DIS, and through this mechanism develop into other national standards.

The standardization of ISO is accomplished by technical committees. Currently, over 130 technical committees have been established. Technical Committee ISO/TC 97 is responsible for developing standards recommendations relating to Computers and Information Processing. The committee's development work is accomplished through the efforts of 14 subcommittees.

One other standards organization has considerable impact on the computer industry of the United States. That is the European Computer Manufacturers Association (ECMA). This body parallels CBEMA of the United States organizationally, but restricts membership in the standards development and approval process to manufacturers only. Although ECMA is not a member of ISO, it is regarded as being a competent standards development body, and its proposals are accepted as a basis for ISO Draft International Standards. In the early 1970s, ECMA was responsible for the development of a proposed standard for the PL/1 language.

1.3 NASA Standardization Activities

Over the past two decades, NASA [4] has invested a large amount of capital in communications and data handling facilities to support space flight projects. To insure effective use of these support facilities, NASA data systems designs are governed by two formally established sets of data system standards. These are the Aerospace Data System Standards and the NASA Planetary Program Data System Standards. These standards define the basic characteristics of the flight/ground interface to assure acceptable support for either deep space or earth-orbiting vehicles by their respective ground-based support complexes. Each places limits on the requirements that may be placed on a NASA tracking network and its associated data processing complex.

The oldest, the Aerospace Data System Standards, was initiated in the early 1960s by the Goddard Space Flight Center to govern the use of the Space-flight Tracking and Data Network (STDN) and its supporting computer complex. The NASA Planetary Program Flight/Ground Data System Standards, established in 1972, governs the use of the Deep Space Network and the Mission Control and Computing Center at the Jet Propulsion Laboratory (JPL).

NASA supports an Intercenter Committee on Automatic Data Processing (ADP) that meets periodically and has representatives from all NASA centers. Its functions are in the areas of acquisition and planning, coordination of standardization, and software standards. The Committee published, in 1979, a document [5] that describes standards for computer resources management.

Some NASA programs carry a standardization activity of their own. Some examples are the End-to-End Information System (EEIS) at JPL that is revising the Planetary Standards. The Applications Data Service (ADS) Standards Program at the Goddard Space Flight Center is working on a set of standards for the Office of Space and Terrestrial Applications (OSTA) of NASA; the Data System Requirements Committee, also at Goddard, maintains the Aerospace Data System Standards.

The NASA End-to-End Data System (NEEDS) program has an activity to coordinate all the standardization activities mentioned above.

SECTION 2

2. DATABASE MANAGEMENT SYSTEMS (DBMS) STANDARDS

At the time when this report was written there were no officially published DBMS Standards by any of the major standardization bodies. However, there is an intense activity going on in the DBMS community and standards organizations to identify interfaces that can be the subject of standardization, to justify the necessity of standards, and to develop standards.

At the present time, the only official DBMS standards that are expected to be published within the next two years are a Data Definition Language (DDL) ANSI standard (see Section 2.2.1.1) and a Data Manipulation Language (DML) ANSI standard for COBOL (see Section 2.2.1.2). All of the other DBMS standardization efforts are at a stage of development that is not expected to produce an official standard in the very near future. However, the publication of the first official DBMS standard will be the first step for further adoption of a similar standard by other standards-setting organizations such as ISO or NBS, and will open the way for further approval of other standards now under development.

In this section, the present status of the DBMS standardization effort will be presented. And an effort will be made to explain what products can be expected from all the DBMS standardization committees, with an estimate of the time frame within which a standard may be issued by the parent organization, and the degree of acceptance or adoption by the user community of the proposal under investigation.

Much of the present DBMS standardization effort stems from the work that the CODASYL organization has been carrying out since 1967, when the CODASYL Database Task Group (DBTG) was founded. This name was retained until they were formally disbanded in 1971.

The first report from the DBTG came out in January 1968 [6] and was entitled "COBOL Extensions to Handle Databases." In October 1969, the first set of language specifications emerged. 1971 contained two major milestones in the life of CODASYL database specifications. The DBTG's April 1971 report [7] was accepted by the parent committee, then called the Programming Language Committee, at a meeting in Washington in May 1971. The acceptance was not entirely unanimous. IBM voiced objections since they had a fairly major investment in their own hierarchical system (IMS), and it was not in their best commercial interest to see a radically different approach move close to standardization.

In May 1971, it was decided that the schema Data Definition Language (DDL), which is to be used by a data administrator for defining a database, should not be a part of COBOL. Therefore, a separate standing committee, called the Data Description Language Committee (DDLC) (see Section 2.1.1), was created to study the problems of database description. The language for specifying the part of a database to be processed by a COBOL program, called the COBOL subschema DDL, and the statements to be added to the COBOL Procedure Division in order to allow a programmer to manipulate the data in a database, called the COBOL Data Manipulation Language (DML), were both formally referred to the Programming Language Committee for consideration as extensions to COBOL.

The Programming Language Committee formed, as a consequence, the Database Language Task Group (DBLTG). Their assignment was to take the DBTG's work and mold it into a form which was suitable for inclusion in the CODASYL _COBOL Journal of Development_. Their first report [8] was approved in 1976 for inclusion as part of CODASYL COBOL and was published in the 1976 CODASYL _COBOL Journal of Development_ [9].

2.1 DBMS Standardization Efforts

Several bodies are currently working on the standardization of database management systems:

1) The CODASYL group;

2) ANSI;

3) International organizations (ISO, ECMA, IFIP);

4) NBS; and

5) Individual federal agency standard bodies.

2.1.1 CODASYL

The CODASYL Data Description Language Committee (DDLC) is a standing committee of the CODASYL organization. Among other functions, its purpose is to provide specifications for the declaration required to establish and maintain database structures [10].

The DDLC publishes its language specifications in a _Data Description Language Journal of Development_ (DDL JOD), and this journal is periodically republished in order to publicize the changes made by the DDLC to its language specifications. Similarly, the CODASYL Programming Language Committee (COBOL) publishes a JOD on COBOL and its database extensions.

The Database Administration Working Group (DBAWG) is a group jointly responsible to the British Computer Society (BCS) and to the CODASYL Data Description Language Committee (DDLC). Historically, it originated as a working group of the BCS Advanced Programming Specialist Group, concerned with the implementation of the CODASYL DBTG proposals [11]. It later changed the emphasis of its work to facilities for data administration.

In December 1973, the DDLC approved a charter to set up a task group on data administration. Because of its previous work, the BCS working group was invited to fill this role, and hence the BCS/CODASYL Data Description Language Committee/Database Administration Working Group (DDLC/DBAWG) came into being. A more detailed history of the DBAWG is given in the June 1975, DBAWG Report [12].

The DBAWG aims to develop techniques for use by data administrators and to produce language specifications where appropriate. The following are some of the activities in which the DBAWG is presently involved:

1) Investigation of the database administrator control function;

2) Investigation of reorganization facilities;

3) Investigation of distributed databases;

4) Support for relational schemas; and

5) Liaison with the BCS Data Dictionary Working Group (ANSI X3H4);

CODASYL neither intends nor considers its proposals to comprise a complete design of a database management system. Rather, its proposals should be viewed as proposed standards for three of the many interfaces between a DBMS and its users and implementors. These three interfaces are [13]:

1) That used by a data administrator to describe the data stored in an on-line database;

2) That used by an application programmer to access and process data in an on-line database; and

3) That which allows data to be allocated to (mapped onto) storage and storage devices.

Many other interfaces exist between a DBMS and its users, for example:

1) That which allows data descriptions to be changed and which causes these changes to be reflected in the database (and its storage);

2) That which allows data recovery for example, checkpointing, dumping and restart; and

3) That which allows utility and service programs to be invoked, for example for editing and printing, for loading the database, for gathering statistics particularly for the use of data, etc.

Some of these and other interfaces are currently being studied by CODASYL.

2.1.2 <u>American National Standards Institute (ANSI)</u>

In the fall of 1972, the American National Standards Institute (ANSI) committee on Computers and Information Processing (X3) through its Standards Planning and Requirements Committee (SPARC), established a Study Group on Database Management Systems with a charter that states that the group will be responsible to:

1) Investigate the subject of database management systems with the objective of determining which, if any, aspects of such systems are, at present, suitable candidates for the development of American National Standards;

2) Provide liaison between the related projects in X3 subcommittees; and

3) Develop and provide United States representation in the international ISO/TC97/SC5W Working Group on Databases.

The study group has three basic products:

1) SD-3s (recommendations for standards, see ANSI document X3/SD-3);

2) Individual reports preparatory to SD-3s; and

3) Standing papers, which constitute the study group's corporate memory and acts as an internal progress measure.

The study group issued an interim report in 1975 [14], a final report in July 1977, and a "Standing Paper 1" [15] in 1980.

"Standing Paper 1" is an organized framework for the approved written work of the ANSI/X3/SPARC Database Systems Study Group. The scope includes material from individuals, task groups, and "The ANSI/X3/SPARC DBMS Framework, Report of the Study Group on Database Management Systems," 1977 [16].

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 responsibility for defining the subschema Data Description Language (DDL) and Data Manipulation Language (DML) specifications to the COBOL committee, referred actions for a subschema DDL and DML specification to the FORTRAN committee, and initiated a committee for DDL specifications.

At present, ANSI has the following groups working on the development of DBMS related standards:

X3H1	Operating Systems Command and Response Language
X3H2	Data Definition Languages
X3H4	Information Resource Dictionary System
X3J3	FORTRAN (and database extension)
X3J4	COBOL (and extensions for database)
X3T5	Open Systems Interconnection

The organization of the ANSI/X3/SPARC Database Systems Study Group includes a Database Architecture Task Group to recommend the development of architectural standards for database systems, and a Relational Database Task Group that is investigating the justifiability of proposing to ANSI/X3 that a standard be developed concerning the relational approach to databases.

2.1.3 International Organizations (ISO, ECMA, IFIP)

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) Defining concepts for conceptual schema languages;

2) Defining or monitoring definition of conceptual languages;

3) Developing a methodology for assessing proposals for conceptual schema languages;

4) Assessing candidate proposals for conceptual schema languages;

5) Defining concepts for conceptual level and user facilities;

6) Defining conceptual level and user facilities;

7) Taking cognizance of and reacting to other database developments as appropriate; and

8) Developing vocabulary for database management systems.

The architecture to be developed by ISO will use as its principal basis the concepts of the interim report that the ANSI/X3/SPARC Study Group on Database Management Systems has issued.

The European Computers Manufacturers Association (ECMA) has chartered a Technical Committee 22 whose task is to develop a standard for database management systems, that is COBOL oriented.

The International Federation for Information Processing (IFIP) has Technical Committee 2 (TC2) on programming, which has a Working Group 2.6 on database management systems. This committee has sponsored several working conferences on various aspects of database management systems. Another TC2 committee (WG 2.8) deals with operating systems command and response languages.

2.1.4 National Bureau of Standards (NBS)

In March 1977, the Federal Information Processing Standards Task Group 24 on Database Management System Standard initiated a study of the need for database standards within the federal government. The voluntary participation from several federal agencies considered the actions of other standard-setting 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.

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

The TG-24 final report [17] issued in August 1979, contains the recommendations for standards and guidelines as well as the assumptions, benefits, and costs considerations used to justify the recommendations.

Other related efforts of the Institute for Computer Sciences and Technology (ICST) have been initiated within a program to develop computer networks protocol standards. The objectives of this program are to (1) make possible distributed computer networks and (2) enable the interconnection of different components selected based on cost, performance, and availability. Some of the protocols product of this program that can be of interest to the DBMS standardization effort are the Distributed Data protocols and the Common Command Language (see Sections 2.4 and 2.5).

2.1.5 Individual Federal Agency Bodies

Among the larger federal agencies participating in FIPS TG-24, several have their own standards-setting function, and some are 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.

2.2 DBMS Standardized Models and Schemas

DBMSs effectively separate the logical and storage structure of data. Thus, users can access data by name and logical interrelationships rather than by location. As a result, the way in which data is physically stored can be modified without affecting program access. Moreover, the same data can be stored by several programs.

Data sharing can be facilitated through using schemas. A schema is the definition, from a particular perspective, of a database. The schema defines entities modeled by the database, attributes of entities, and relationships between entities. The schema specifies a class of permissible database states and permissible transitions between database states.

The 1977 ANSI/X3/SPARC DBMS Framework, Report of the Study Group on Database Management Systems, specified three types of schemas:

1) External Schema: an external schema is a logical description of a portion of an enterprise. It describes the user or user program view of data.

2) Conceptual Schema: the conceptual schema is the comprehensive, base-level, logical description of the environment in which an enterprise exists. It should be free of both physical structure and application system considerations. It may describe entities and attributes that are currently not retained in physical storage.

3) Internal Schema: an internal schema describes data modeling an enterprise as it is physically stored. It includes all aspects of the environment in which the database is to reside (device types, access methods, etc.).

These three schemas can be viewed as layers through which a request passes in actually accessing data managed by a DBMS. Consequently, processing a request requires mapping the external schema to the conceptual schema which, in turn, is mapped to the internal schema.

DBMSs can be categorized in terms of the data model which is supported and the data language provided for interacting with this data model. A data model defines the acceptable types of data structures. The three basic data models discussed below are logically based on the representation of data as record types; the network and hierarchical models also support explicit system maintenance of record type interrelationships.

Developments in the DBMS field, unlike those in programming languages, have brought about no significant degree of de facto standardization. For the three commonly recognized data models (the hierarchical, the network or CODASYL model, and the relational model) industry statistics show that about 30 percent of installations use hierarchical systems, 25 percent use CODASYL systems, and 45 percent use non-CODASYL network systems [18]. There are not yet a significant number of commercial installations of relational systems. But there is a widespread interest in this approach, which is recognized by the major standards organizations in providing standards.

2.2.1 Network Model

The network data model is widely accepted (see Section 2.2) and it corresponds to the DBTG proposal [7 and 19]. In fact, the DBTG data model is the network data model described in [20] with two restrictions:

1) All relationships of data items must be 1:N (this corresponds to the DBTG rule that no record occurrence can participate in more than one set occurrence of the same type), i.e., complicated N:M relationships are prohibited.

2) No recursive links are allowed (this corresponds to the DBTG rule that a record type cannot be both owner and member at the same time within a set type).

There is also some difference in terminology. For instance, links are called set types by the DBTG. In addition, the DBTG proposal includes some other features, e.g., aggregates, multimember set types, etc. However, the DBTG proposal is, in essence, the network data model.

The DBTG proposal is widely accepted by the database community and has been adopted as a base of work by the major standardization organizations (ANSI, ISO). It will be the first DBMS standard to be published (see below), and it is expected that other standards organizations (other than ANSI) will follow and present a similar CODASYL-based standard for publication.

2.2.1.1 Data Description Language (DDL). In 1978, the CODASYL Data Description Language Committee (DDLC) published its last Journal of Development [19] where it presented its proposal for a Data Description Language. The 1978 Journal of Development is the basis of work for the ANSI X3H2 committee on Data Description. ANSI X3H2 made some modifications to the CODASYL work, and at a very recent meeting the proposal was turned down by the committee. The proposal failed because of the lack of a COBOL facility to access a database on which to base some of the DDL standard work. As a consequence, at the time when this survey was published, X3H2 is changing its charter to include the semantics of operation of the database facility. Although, it is premature to predict when the proposal for a standard will be ready, an official DDL standard is not expected to be available before 1983.

The DDLC has defined two DDLs:

1) The most important DDL is that used to declare a schema. The schema is the logical description of all the data stored in a database. The schema DDL is considered to be language (i.e., COBOL, FORTRAN, PL/1) independent and it does not include the definition of physical storage media. This language will normally be used by a "data

administrator" that is a specialist responsible for ensuring that the database meets, preferably optionally, the requirements for data placed on it by its users. This language is specified in the 1978 DDLC Journal of Development [19].

2) To a user of a DBMS, and by "user" CODASYL normally means an application programmer using COBOL, the only DDL is that used to declare a subschema. A subschema is the description of a database or a part of a database as the user himself sees the data, and is thus dependent upon which programming language (i.e., COBOL, FORTRAN, PL/1) the user uses and upon the users application.

The International Standards Organization (ISO) is not currently working on a DDL. However, it is expected that it will base its future work on the ANSI CODASYL-type proposal.

2.2.1.2 Data Manipulation Language (DML). The data in a database or part of a database described by a subschema is manipulated by obeying a program (or, in CODASYL terms a "run unit," because asynchronous manipulation of the database by many users is envisaged) written using a mixture of standard high-level language (i.e., COBOL, FORTRAN, PL/1) commands and Data Manipulation Language (DML) functions. The DML is independent of the high-level language, and thus is not an element of its syntax; the high-level language is described as "hosting" the DML. The DML is also dependent upon the subschema DDL in the sense that the "run unit" using the DML may only operate on data declared in a subschema.

The CODASYL DML proposal was developed by the Database Language Task Group (DBLTG) of the CODASYL Programming Language Committee (COBOL), and it is specified in the CODASYL COBOL Journal of Development.

The ANSI X3J4 committee on Data Manipulation Language works on the specifications provided by CODASYL. This committee does not do any development work (as opposed to X3H2 on DDL) and will try to coordinate with X3H2 when the proposal is sent out for public review. A standard will not be available before 1983.

2.2.1.3 Storage Description Language (DSDL). A Data Storage Description Language (DSDL) defines how data described in a schema may be organized in terms of an operating system- and device-independent storage environment. Such a description is known as a storage schema. A storage schema has no effect on the results of application programs but only affects their performance.

A storage schema is created for a schema by the data administrator, application programmers, and subschema writers need not be aware of the design of the storage schema. The concept of separate schema and storage schema achieves the separation of the logical description of the entire database from the storage description.

The language is intended to be used with the schema DDL specified in the CODASYL DDLC Journal of Development.

The DSDL has only been specified in a draft form [21] as an appendix to the 1978 DDL JOD [19]. There are no clues as to the time frame within which this draft proposal may become a standard, since ANSI has not as yet taken any actions concerning a DSDL.

2.2.1.4 Fortran Database Facilities. The schema Data Description Language (see Section 2.2.1.1) has been designed to be independent of the high-level language (e.g., COBOL, FORTRAN, PL/1). The subschema Data Description Language, however, is dependent upon the host high-level language. The Data Manipulation Language (DML) is said to be "hosted" by the high-level language.

The subschema Data Description Language and the DML are, thus, described within the environment of the specifications of each high-level language (i.e., the DML proposed by CODASYL and intended to be used in a COBOL environment is specified in the CODASYL COBOL Journal of Development).

Up to the present, there is only one other language (other than COBOL) for which actions have been taken to specify database facilities intended to, eventually, be part of that language's standard. That language is FORTRAN.

The CODASYL FORTRAN Database Language Committee (FDBLC) published, in January 1980 [22], a Journal of Development with the specifications of the FORTRAN Database Facilities.

The ANSI X3J3 Committee on FORTRAN has taken the CODASYL proposal as the base for their work. However, it is expected that the earliest date by which a standard will be ready is 1985.

2.2.2 Relational Model

The Relational Task Group (RTG) of the ANSI/X3/SPARC Database Systems Study Group is currently working on the compilation of a feature analysis catalogue of relational systems to determine what constitutes relational capabilities. It is also investigating the justification for, and feasibility and timeliness of, its relational standardization activity. The products of the RTG will be a preliminary and a final report. The preliminary report will present relational concepts and terms and the feature catalog. The final report will contain an SD-3 (recommendation for a standard) and technical requirements for a relational standard [23].

There are two principal alternative approaches to achieving relational capabilities in standard specifications:

1) Incorporating relational features in standards currently being developed (e.g., by X3H2, X3J3 and X3J4); and

2) Developing a relational standard independent of other standards.

There are indications that the second approach will be taken.

Note that this work does not yet involve the development of a standard, although it most likely will evolve towards the specification of one.

The CODASYL Data Description Language Committee is also investigating relational schemas.

2.2.3 Hierarchical Model

There is currently no significant work being done towards the specification of a standard for the Hierarchical Model.

2.3 Data Dictionary (DD)

In the proposed architecture of the ANSI/X3/SPARC Database Systems Study Group [15], the Data Dictionary (DD) is viewed as a repository of information about the entire processing environment and its usage (i.e., "meta-data"). In this context, the DD is responsible for the following general functions:

1) Providing a source of meta-data for documentation of the environment (i.e., processes, users, data characteristics, transactions, and outputs); and

2) Providing a source of current information to support system control functions (i.e., access and function tables, validation rules/tables, access statistics, etc.).

As a result of a recommendation made by a task group of the ANSI Database Systems Study Group, ANSI Committee X3H4 was formed in July 1980 to develop a standard for an Information

Resource Dictionary System. The proposal indicates a target date of 42 months after inception of the technical committee to approve a national standard. See also reference [24].

The National Bureau of Standards (NBS) will publish the specification of a Federal Dictionary System Standard as a Federal Information Processing Standard (FIPS), so that it can be used with General Services Administration regulations for federal procurement purposes.

The federal DDS standard will not require agencies to use a DDS in all applications. However, wherever an agency decides to implement DDS software or services, the federal DDS standard specification shall be used as the basis for procurement action.

In September 1980, the NBS issued a document [25] that describes the project which NBS has initiated to develop a federal standard for Data Dictionary Systems. It also provides information about the use and benefits of Data Dictionary Systems.

NBS will develop a Functional Requirements Report early in Fiscal Year 1981. The Functional Specification will be distributed for public review early in FY82. It is expected that further development will proceed so that official review of a candidate standard takes place early in FY83. Then, a recommended standard would be forwarded for Secretary of Commerce approval by the close of FY83.

One of the NASA standardization programs, the OSTA/ADS Standards Program has an activity to develop catalog standards for OSTA/ADS. A preliminary report on the requirements for standards for catalogs, directories and dictionaries appears in the soon-to-be-published proceedings of the May 27-29, 1981 OSTA/ADS Data Systems Standards Workshop at the Goddard Space Flight Center.

2.4 End-User Facilities

2.4.1 CODASYL

The End-User Facility Task Group (EUFTG) of the CODASYL organization was formed in 1972. It has been working since then on the development of specifications for an End-User Facility (EUF) which will enable user-computer personnel to access information stored in computerized files. In 1978, the group published a status report on its activities [26], and it is currently working on an _End-User Facilities Journal of Development_ which will contain the specification of a forms-oriented End-User Facility. The forms approach was considered the most natural interface between an end user and data because a large number of users employ forms (e.g., purchase order forms, expense report forms, etc.) or versions of forms (e.g., reports, memos, etc.) in their daily work activities. The committee has taken the approach to initially specify a generalized forms creation and processing capability, as the primary orientation for its work.

2.4.2 ANSI/ISO

In 1977, ANSI/X3 established the Study Group on Distributed Systems (DISY). DISY was later reorganized into the SPARC Open Systems Interconnection Committee (OSIC). In March 1980, OSIC was designated a Technical Committee of ANSI/X3, titled X3T5 Open Systems Interconnection.

Beginning in 1978, ANSI Committees DISY and OSIC/X3T5 have been participating with Subcommittee 97/16, Open Systems Interconnection of ISO, in developing the Reference Model of Open System Interconnection [27] which represents the basis for coordinated standards development in this area. The Reference Model describes the basis in terms of seven hierarchical layers. The seventh layer (Application layer) would include the End-User Facilities as part of its protocols.

The seven-layer model is now generally accepted by ISO, European Computer Manufacturers Association (ECMA), International Telegraph and Telephone Consultative Committee (CCITT), and many national organizations, including ANSI. It will become an ISO standard in the very near future.

Present principal activities of ISO and ANSI (and others) dealing with the sixth (Presentation) and seventh (Application) layers concern the following aspects:

1) Virtual Terminal Protocols (VTP)

2) Virtual File Service (VFS)

3) Job Transfer and Manipulation Protocols (JTMP)

4) Upper layers architecture

5) Management Protocols

The VTP activity is defining the protocols that are most useful for operations where at least one session partner is a terminal. The VFS work deals with the protocols most useful in the transfer of quantities of information. The JTMP group objective is to provide protocols which support the control and monitoring of multisite job activities; this group is inactive at the present time. The architecture group is coordinating the above efforts so that a consistent set of protocols is developed for each of the two layers. Finally, the management protocols group deals with issues such as authorization, security, error handling, resource availability, reporting and accounting, audit trails, journaling, etc. [28] through [32].

2.4.3 NBS

There are two relevant activities going on at NBS:

1) Computer Sciences Corporation (CSC) is currently working on a Query Language for the NBS Task Group on Database Management Systems (TG24). This Query Language is not intended to have update capabilities.

2) The Institute for Computer Sciences and Technology (ICST) of NBS has initiated a program to develop computer network protocol standards. The Network Protocol Standards Program will produce a series of standards and protocols of which two may be relevant to the upper layers of the Reference Model of Open Systems Interconnection:

   a) A Common Command Language (CCL) is being developed that will provide a user interface to a Data Transfer Protocol which will permit the sending of files of buffers between computer systems on a network. CCL has some query capabilities to retrieve records from a single file according to a single key. This is a major difference from a query language in a DBMS environment, which does handle relationships between files. A draft report on the CCL is available [33]. The final report will be available by 1982.

   b) Distributed Data Protocols. This set of protocols will facilitate user access to distributed data in a networking environment. The final report will be available by 1985.

2.5 Distributed Databases

There is some effort in the major standardization bodies concerning the issue of standards for distributed databases. However, this work is not expected, at present, to produce any standards in the very near future.

In its Network Protocol Standards Program, NBS is addressing this issue as was described in Section 2.4.3.

In January 1977, the CODASYL Systems Committee undertook a task examining the implications of database technology in a distributed environment. The committee is focusing its attention primarily on the impact of extending database management techniques to distributed processing environments. The scope of this effort is the establishment of a framework for examining these issues, and the formation of baseline concepts and guidelines which will support the continuing development of distributed database management technology. The result of this work will be published as a CODASYL Systems Committee Technical Report. An interim report and a commentary are in [34] and [35].

The ANSI/X3/SPARC Database Study Group is also investigating distributed databases. There is, at present, no ANSI Technical Committee developing a standard. However, the Study Group has a Distributed System Task Group that has produced a proposal for an architecture of a distributed information resource [36] and a discussion of the role of a global DDL, global DML, and global Query Language in the context of a distributed database management system environment [37].

SECTION 3

3. FILE STRUCTURE FOR INFORMATION INTERCHANGE AND STORAGE

3.1 NASA Format Standards

There are not, at present, any Information Interchange or storage standards that are universally accepted within NASA. Following the policy that has governed the design of NASA missions up to the present, each mission has required the development of mission-unique designs of information systems.

There is a significant amount of effort within NASA to arrive at a unified set of standards for information processing (see also Section 1.3). Some examples of NASA programs for which there is an activity in the investigation of standards are the NASA End-to-End System (NEEDS) program sponsored by the Office of Aeronautics and Space Technology (OAST), the End-to-End Information System (EEIS) at JPL and the Applications Data Service (ADS) Standards program at Goddard. Some of these programs are producing some specifications that are being adopted by more than one mission, and they are good *de facto* candidates for official NASA multimission standards. In particular, the Standard Formatted Data Unit (SFDU), (see Section 3.1.2) has become a very important candidate for a data storage and interchange format standard.

Another good source of candidates for standards are some of the major projects (i.e., Landsat) for which certain mission standards have been developed. A few are described in the following sections.

The scope of this study will not include standards relevant to the lower layers of the Reference Model for Open Systems Interconnection [27]. Specifically, communication formats are not included.

3.1.1 Source Packet Format

The NASA End-to-End Data System (NEEDS) program, sponsored by the Office of Aeronautics and Space Technology (OAST), has adopted the concept of packet telemetry for the architecture of the information system that it is proposing.

The packet telemetry concept was originally presented in [38], when it was called the Instrument Telemetry Packet (ITP) concept. It represents an alternative to the conventional multiplexed telemetry frame approach for acquiring spaceborne instrument data.

A format has been designed to support this concept, called the "Source Packet Format." It is a level 6 protocol of the Reference Model for Open Systems Interconnection, and it is described in [39].

The EEIS program at JPL has also adopted the packet telemetry concept for Deep Space missions. Reference [40] presents a set of general guidelines for the implementation of packet telemetry, followed by specific standards which are to be observed by families of flight projects. The guidelines are fully aligned with the NEEDS effort [39].

The Source Packet format was not designed for storage, and it would have to be modified if that was its intended use (the mission ID field within the primary header has 8 bits that would allow for 256 mission ID numbers only).

3.1.2 Standard Formatted Data Unit (SFDU)

The SFDU defines a family of standard message structures to be implemented for electronic data exchanges which occur within ground facilities which are components of the Multimission End-to-End Information System (MMEEIS) at JPL. The SFDU has become an important candidate for a NASA-wide format standard and there is an effort to turn it into an international standard for the exchange of space-derived information.

The SFDU is described in [41] and is compatible with the work of ANSI committee X3S33 (Task Group 3) Project 281 on Code Independent Message Heading Formats (see [42]).

3.1.3 Image Data Standards

3.1.3.1 VICAR. The Video Information Communication and Retrieval System (VICAR) is operational in several institutions inside and outside of NASA in a variety of applications. The wide acceptance of the NASA VICAR format makes it a good candidate for a standard.

This format allows unlimited amounts of ancilliary data per image. However, it has only been used to date as a processing and storage (not transmittal) mechanism and is currently quite machine dependent in pixel storage codes. Further, it is not designed for use with nonuser image data (feature vectors, etc.) and retains only limited flexibility for distinction between tape and image formats. The VICAR format is described in [43] and [44].

3.1.3.2 Multimission Imaging EDR (Experimental Data Record) Format. The JPL Regional Planetary Image Facility has been developing software to recover the imaging data from all past JPL missions that is on the archived Master Data Record (MDR), Experimental Data Record (EDR) and RDR tapes. Plans are to recover this data, reformat it into a common Multimission Image EDR, and write it onto 9-track, 1600 bpi tapes during FY81. The concurrent part of this effort is to get Galileo and all future missions to produce the imaging EDR in this multimission format.

The Multimission Imaging EDR format proposal will store image data in a VICAR compatible format, but, unlike VICAR, it was not designed for VICAR systems only. This format is described in [45].

3.1.3.3 Computer Compatible Tape (CCT) Format. The Landsat-D CCT Standards Committee has been examining Computer Compatible Tape (CCT) formats and format philosophies, with the objective of establishing some standards for CCTs which would promote information exchange among remote sensing data users and would allow data from a variety of sources to be used for a given application. Format requirements were collected from members and a draft document [46] of these requirements was prepared in 1978 by the Canada Center for Remote Sensing (CCRS).

In 1979, NASA, with the assistance of CCRS, published a document [47] to describe and define the tape format standardization approach recommended by the committee on CCT standardization. This approach applies to all types of remote sensing data user tapes. The purpose of the proposal is to address user tapes (whose data are extracted from the production/archive tape and reformatted to provide a product more suitable for the user), as opposed to production tapes (used as a digital archive) that may have additional system-imposed requirements that were not addressed by the committee.

The key to the CCRS approach is a concept which, in the document, is referred to as a superstructure (a combination of precisely defined records and a method of employing them) which, when combined with any particular tape format, provides access to the data of that format without requiring specific knowledge of the particular format specifications.

Two standard CCT formats have been developed to date that include the CCT Family of Tape Formats Superstructure conventions and conform to the design standards for CCT format as established by the Landsat-D CCT Standards committee [47]:

1) The Landsat-D Computer Compatible Tape format described in [48] defines the data format for Computer Compatible Tapes that are produced by the Data Management System (DMS) at the Goddard Space Flight Center. These tapes serve as input media to the Earth Resources Observation System (EROS) Data Center (EDC) of the United States Department of the Interior.

2) An interagency committee Spatial Data Transfer Committee (SDTC) of the Government of Canada has published a standard format for the interchange of spatially encoded data in the form of polygon chains [49].

3.2 External Standards

3.2.1 Labels and File Structure for Information Interchange

These conventions are transport protocols of the Fourth (Transport) layer of the Reference Model for Open Systems Interconnection [27].

3.2.1.1 Information Interchange Data Descriptive File. The ANSI X3L5 subcommittee has issued for public review and comment, a working paper [50] on a proposal for a standard that establishes media- and machine-independent formats to facilitate the interchange of information between computing systems. It is intended for use with physical recording media as well as communications media. It is not designed as a record format for retention within the files of any specific installation, and thus processing efficiency is not a consideration.

The standard accommodates data elements in a hierarchical interchange structure. It can accept the commonly used data structures, many of them directly and others in equivalent forms. While the standard may find use in interchanging some CODASYL databases, it was not intended to be readily usable for this purpose.

The International Standards Organization (ISO/TC97/SC15 N61) established Working Group 3 (WG3) to recommend scopes of and priorities for International Standards to be developed for the Interchangeable IRV Coded Data Files project (Programme 97.15.6). The project coincided with the completion of the draft proposed ANSI standard on the equivalent topic. ISO is presently investigating [51] the ANSI proposal [50].

3.2.1.2 Magnetic Tape Labels and File Structure. ANSI standard X3.27 establishes a standard for information interchange utilizing magnetic tape, by providing magnetically recorded labels to identify and structure files, and by providing a standard structure for the blocks containing the records that constitute a file [2].

3.2.1.3 Magnetic Tape Cassette and Cartridge Labels. The International Standards Organization (ISO/TC 97's) project number 97.15.3 issued standard ISO 4341-78 on this topic.

The European Computer Manufacturers Association (ECMA) has standard 41 on the same subject.

ANSI committee X3/L5 is also working in this standard (project 217). It will issue a proposal (very similar or identical to the ISO standard) in the very near future.

3.2.1.4 Flexible Disk Labels and File Structure. ISO has a proposal for a standard (DIS 6863, project 97.15.5).

The European Computers Manufacturers Association (ECMA) has adopted this standard as standard #58.

ANSI is investigating ISO's proposal (ANSI Committee X3/L5, project 272).

3.2.1.5 Volume Labels on Magnetic Media. ANSI is working on a proposal for a standard. It is expected that the proposal will be issued for public review and comment before the end of 1980.

3.2.2 Data Element Representation

3.2.2.1 Code for Information Interchange. ANSI standard X3.4-1977 denotes the coded character set to be used for the general interchange of information among information processing systems, communication systems, and associated equipment. The notation ASCII should ordinarily be taken to mean the code prescribed by the latest edition of this standard.

3.2.2.2 Representation for Calender Date and Ordinal Date for Information Interchange. ANSI standard X3.30 and Federal Information Processing Standard (FIPS) 4 specify standard means of representing calendar date (year, month of year, and day of month) and ordinal date (year and day of year) to facilitate interchange of data among data systems.

3.2.2.3 Structure for the Identification of the Counties of the United States for Information Interchange. ANSI standard X3.31 and FIPS 6-2 identify and provide a 3-digit numeric code structure for counties and county equivalents of the states of the United States, including the District of Columbia [2].

3.2.2.4 Graphic Representation of the Control Characters of American National Standard Code for Information Interchange. ANSI standard X3.32-1973 and FIPS 36 provide for a graphical representation of the control characters of columns 0 and 1, and for the characters Space and Delete, of American National Standard Code for Information Interchange X3.4-1977 (ASCII), containing alternative pictorial and alphanumeric sets of representations for use on display devices where graphical representations of these normally nonprinting characters are required [2].

3.2.2.5 Identification of States of the United States (Including the District of Columbia) for Information Interchange. ANSI standard X3.38-1972 (R 1977) and FIPS 5 provide 2-digit numeric codes and 2-character alphabetic abbreviations for states of the United States and District of Columbia [2].

3.2.2.6 Code Extension Techniques for Use With the 7-bit Coded Character Set for American National Standard Code for Information Interchange. ANSI standard X3.41-1974 and FIPS 35 specify standard procedures for augmenting the 128-character repertory of American National Standard Code for Information Interchange X3.4-1977 (ASCII), with additional graphics or control functions. It provides a code extension structure that will also accommodate the need for 8-bit codes for general information interchange of which ASCII is a subset [2].

3.2.2.7 Representation of Numeric Values in Character Strings for Information Interchange. ANSI standard X3.42-1975 specifies the syntax as the elements of three sets of character sets, of character strings, which are decimal positional representations of numeric values, that is, numeric representations [2].

3.2.2.8 Representations of Local Time of the Day for Information Interchange. ANSI X3.43-1977 is designed to establish uniform time representations based upon both the 12- and 24-hour timekeeping systems. It provides a means for representing local time of the day in digital form for the purpose of interchanging information among data systems [2].

3.2.2.9 Structure for the Identification of Named Populated Places and Related Entities of the States of the United States for Information Interchange. ANSI X3.47-1977 provides the structure for an unambiguous, 5-digit code for named populated cities, towns, villages, and similar communities, and several categories of named entities similar to these in one or more important respects.

3.2.2.10 Representations for United States Customary, SI and Other Units To Be Used in Systems With Limited Character Sets. ANSI X3.50-1976 provides representations for units of weights and measures to be used in data interchange systems with limited graphic character set capabilities. The representations apply to names of United States customary and other internationally recognized units and their decimal multiples and submultiples [2].

3.2.2.11 Representations of Universal Time, Local Time Differentials, and United States Time Zone References for Information Interchange. ANSI X3.51-1975 provides standard means for representing universal time, local time differentials, and United States time zone references to facilitate interchange of data among data systems [2].

3.2.2.12 <u>Representation of Geographic Point Locations for Information Interchange</u>. ANSI X3.61 establishes uniform formats for geographic point location data and provides a means for representing these data in digital form for the purpose of interchanging information among data systems [2].

3.2.2.13 <u>Additional Controls for Use With the American National Standard Code for Information Interchange</u>. ANSI X3.64-1979 has as its primary purpose to provide a general set of controls to accommodate the foreseeable needs in the following diverse information interchange applications: (1) interactive terminals of the cathode ray tube type; (2) interactive terminals of the printer type; (3) line printers; (4) microfilm printers; (5) software usage; (6) form filling; (7) composition imaging (for example, typesetting); (8) word processing; (9) input-output devices with auxiliary devices; and (10) buffered and unbuffered devices [2].

3.2.2.14 <u>Guidelines for the Organization and Representation of Data Elements for Data Interchange</u>. ISO committee TC97/SC14 (Representation of Data Elements) has a not-yet approved first draft proposal which is a guide to assist users and designers of data processing systems to identify, define, and represent units of data to be interchanged along with the necessary structures for organizing these data.

3.2.3 External Image Standards

At the present time, there are no officially published or generally accepted standard image formats. There are some standardization efforts being done by the NBS [52], but it is not expected that they will produce a standard in the very near future. A list of some candidates for a standard appears below.

3.2.3.1 <u>The NATO RSG-4/SGIP Tape Format.</u> Under the NATO Defense Research Group (DRG), the AC/243 (Panel III) Research Study Group 4 (RSG4) was formed to coordinate participation among participating governments in the field of image pattern recognition. The Subgroup on Image Processing (SGIP) issued in February 1976, the specifications for a tape format to be used for transferring digital imagery from one installation to another [53].

The NATO format has been internationally coordinated and is the only one among other candidate formats (NASA VICAR, NASA CCT) that makes provision for transmittal of nonimagery data, real and complex-valued imagery, and negative value image points. The concepts of image and tape format are completely decoupled and treated as separate issues. This allows great flexibility including very long image lines with simultaneous limits on maximum tape buffer length. However, this format places limits on ancilliary data (none in an image line and a maximum of only 4000 characters per image header in free format) which severely limits its utility in NASA type databases.

3.2.3.2 General Specifications for a Tape Format (IEEE Panel on Biomedical Pattern Recognition). A task force on Databases and Portable Software was formed in 1975, within the Biomedical Pattern Recognition Subcommittee of the Machine Intelligence and Pattern Analysis Committee of the IEEE Computer Society. The group has issued some very loose general specifications for a tape format [54].

3.2.3.3 Initial Graphics Exchange Specification (IGES). The Initial Graphics Exchange Specification (IGES) is the product of a program sponsored by the Air Force, Army, Navy and NASA, and coordinated by the NBS. IGES is an attempt at providing a specification for the exchange of drawing and geometry data between Computer Aided Design/Computer Aided Maufacturing (CAD/CAM) systems.

ANSI subcommittee Y.14.26 (Computer Aided Preparation of Product Definition Data) voted on May 1980, to issue IGES as the first three parts of a 5-part proposed standard. IGES is described in [55].

3.2.4 Storage Media Standards

The following are ANSI standards, some of which have been adopted by NBS as FIPS and/or by the Department of Defense (DOD), as indicated in each.

Although the relevance of these standards to DBMSs is questionable, they are included here for completeness.

3.2.4.1 Magnetic Disks

3.2.4.1.1 Unrecorded Magnetic Six-Disk Pack (General, Physical and Magnetic Characteristics). ANSI X3.46-1974. Specifies the general, physical, and magnetic requirements for interchangeability of the six-disk pack between disk storage drives and associated information processing systems.

3.2.4.1.2 Unrecorded Single-Disk Cartridge (Front Loading, 2200 bpi), General, Physical, and Magetic Requirements. ANSI X3.52-1976, (DOD). Provides the general, physical, and magnetic requirements for interchangeability of the magnetic single-disk cartridge (front loading), as required to achieve unrecorded cartridge interchange between disk storage and associated information processing systems.

3.2.4.1.3 Unrecorded 11-Disk Pack, General, Physical and Magnetic Requirements. ANSI X3.58-1977. Specifies the general, physical, and magnetic characteristics required for the physical interchange of magnetic 11-disk packs for use in electronic data processing systems.

3.2.4.2 Magnetic Tape

3.2.4.2.1 Recorded Magnetic Tape for Information Interchange (200 CPI, NRZI). ANSI X3.14-1973. Provides specifications for format and recording for a 1/2-inch, 9-track magnetic tape to be used for information interchange among information processing systems, communication systems, and associated equipment utilizing the ASCII, X3.4-1977. This standard deals solely with recording on magnetic tape and supports and complements American National Standard Unrecorded Magnetic Tape for Information Interchange (9-track 200 and 800 CPI, NRZI, and 1600 CPI, PE), X3.40 1976.

3.2.4.2.2 Recorded Magnetic Tape for Information Interchange (800 CPI, NRZI). ANSI X3.22-1973, (FIPS 3.1). Provides specifications for format and recording for 1/2-inch, 9-track magnetic tape to be used for information interchange among information processing systems, communication systems, and associated equipment utilizing the ASCII X3.4-1977. This standard deals solely with recording on magnetic tape and supports and complements American National Standard Unrecorded Magnetic Tape for Information Interchange (9-track 200 and 800 CPI, NRZI, and 1600 CPI, PE), X3.40-1976.

3.2.4.2.3 Magnetic Tape Labels and File Structure for Information Interchange. ANSI X3.27-1978. Establishes a standard for information interchange utilizing magnetic tape, by providing magnetically recorded labels to identify and structure files, and by providing a standard structure for the blocks containing the records that constitute a file.

3.2.4.2.4 Recorded Magnetic Tape for Information Interchange (1600 CPI, PE). ANSI X3.39-1973, (FIPS 25). Provides specifications for format and recording for a 1/2-inch, 9-track magnetic tape to be used for information interchange among information processing systems, communication systems, and associated equipment utilizing the ASCII X3.4 1977. It deals solely with recording on magnetic tape and supports and complements American National Standard X3.40-1976.

3.2.4.2.5 Unrecorded Magnetic Tape for Information Interchange (9-track 200 and 800 CPI, NRZI, and 1600 CPI, PE). ANSI X3.40-1976. Provides specifications for 1/2-inch wide unrecorded magnetic tape and reels to permit mechanical and magnetic interchangeability of tape between information processing systems, communication systems, and associated equipment utilizing the ASCII X3.4-1077. This standard deals solely with magnetic tape for digital recording and supports and complements American National Standards on unrecorded magnetic tape for information interchange.

3.2.4.2.6 Magnetic Tape Cassettes for Information Interchange (3.810-mm (0.150-inch) Tape at 32 bpmm (800 bpi), PE). ANSI X3.48-1977. Covers specifications and requirements for a 3.810-mm (0.150-inch) magnetic tape cassette to provide data interchange and physical interchangeability between information processing systems utilizing the ASCII X3.4-1977 and amendments thereto.

3.2.4.2.7 Recorded Magnetic Tape for Information Interchange (6250 CPI, Group Coded Recording). ANSI X3.54-1976. Provides format and recording specifications for 1/2-inch, 9-track magnetic tape to be used for information interchange, information processing systems, communication systems, and associated equipment utilizing the ACSII X3.4-1977.

3.2.4.2.8 Unrecorded Magnetic Tape Cartridge for Information Interchange (0.250-inch, 1600 bpi, Phase Encoded). ANSI X3.55-1977. Presents the minimum requirements for the mechanical and magnetic interchangeability of a 0.250-inch (6.30-mm) wide magnetic tape cartridge among information processing systems, communications systems, and associated equipment. This standard refers solely to a magnetic tape cartridge for digital recording and supports and complements ANSI X3.56-1977.

3.2.4.2.9 Recorded Magnetic Tape Cartridge for Information Interchange (4-track, 0.250-inch, 6.30-mm, 1600 bpi, 63 bpmm, Phase Encoded). ANSI X3.56-1977. Provides the technique and format specifications for recording the code of ANSI X3.4-1977, Code for Information Interchange (ASCII), on a 0.250-inch (6.30-mm) wide, 4-track, magnetic tape in a cartridge at 1600 bpi, using phase encoding for information interchange among information processing systems, communication systems, and related equipment. This standard supports and complements ANSI X3.55-1977.

3.2.4.2.10 One-Half Inch (12.7-mm) Magnetic Tape Reel for Computer Use (Requirements for Interchange). ANSI/EIA RS 352-1968 (R1978). Covers those dimensions of reels considered essential for their successful interchange between equipment designed for use with 1/2-inch wide magnetic tape in computer applications.

3.2.4.3 Paper Tape

3.2.4.3.1 Perforated Tape Code for Information Interchange. ANSI X3.6-1965 (R1973), (FIPS 2). Specifies the representation of the ACSII, X34-1977, in perforated tape and similarly encoded media used for interchange of information among equipment such as office machines and information processing and communications apparatus.

3.2.4.3.2 One-inch Perforated Paper Tape for Information Interchange. ANSI X3.18-1974. Covers the physical dimensions of the paper tape which is to be perforated with fully punched round holes.

3.2.4.3.3 Eleven-sixteenths Inch Perforated Paper Tape for Information Interchange. ANSI X3.19-1974. This standard covers specifications for eleven-sixteenths inch tape similar to those provided for one-inch tape specified in ANSI X3.18-1974.

3.2.4.3.4 Take-up Reels for One-inch Perforated Tape for Information Interchange. ANSI X3.20-1967 (R1974), (FIPS 27) (DOD). Covers the physical dimensions of take-up (or storage) reels, with either fixed or separable flanges.

3.2.4.3.5 Specification for Properties of Unpunched Oiled Paper Perforator Tape. ANSI X3.29-1971. Describes the dimensions of unpunched paper perforator tape and the properties of the paper from which it is made. This tape is intended to be used as a perforated tape input/output medium for information interchange among information processing systems, communication systems, and associated equipment.

3.2.4.3.6 Interchange Rolls of Perforated Tape for Information Interchange. ANSI X3.34-1972, (DOD). Provides definitions and dimensions for interchange roll, length of tape in interchange roll, directional markers on tape, and length of header and trailer.

3.2.4.4 Punched Cards

3.2.4.4.1 Specification for General Purpose Paper Cards for Information Processing. ANSI X3.11-1969, (DOD). Specifies the dimensions, quality of paper, and test methods of 7-3/8 inch-length cards used for information processing. This standard is intended to apply to general purpose cards in which the primary method of recording information is by punched holes.

3.2.4.4.2 Rectangular Holes in Twelve-row Punched Cards. ANSI X3.21-1967, (FIPS 13). Specifies the size and location of rectangular holes in 12-row, 3-1/4 inch-wide punched cards.

3.2.4.4.3 Hollerith Punched Card Code. ANSI X3.26-1970, (FIPS 14). This standard specifies 256 hole patterns in 12-row punched cards. Hole patterns are assigned to the 128 characters of the ASCII X3.4-1977, and to 128 additional characters for use in 8-bit coded system.

SECTION 4

4. SOFTWARE DESIGN AND DOCUMENTATION STANDARDS

4.1 NASA Standards

The only NASA-wide set of software guidelines in existence appears in the NASA Software Management Guideline [56] of the Computer Resources Management document [5] that the Inter-center Committee on Automatic Data Processing (see Section 1.3) has issued.

Some major NASA facilities have their own set of Software Design and Documentation Standards (see also Section 1.3). For example, the Goddard Space Flight Center [57] and JPL have a draft Standard Practice [58 and 59] that provides a detailed description and definition of the software life-cycle process, the implementation methodology, and management policies.

4.2 External Standards

The national standards organizations have not issued a unified standard that covers most aspects of the software life-cycle. There is, however, a set of standards that covers some specific aspects of the life-cycle. They are described below.

The NBS has issued a guideline [60] that explains alternative software capbilities and recommended development practices for database applications, with specific advice on applications planning and management, and on software selection. The guideline is intended as a technical primer on basic reference for federal managers and application analysts who are responsible for computer applications and associated software and project decisions. Its use is encouraged for planning and management, but is not mandatory.

4.2.1 Guidelines for Documentation of Computer Programs and Automated Data Systems

National Bureau of Standards/Federal Information Processing Standards Publication 38 (NBS-FIPS-PUB 38), February 15, 1976, provides basic guidance for the preparation of ten document types that are used in the development of computer software. It can be used as a checklist for the planning and evaluation of software documentation practices.

4.2.2 Guidelines for Documentation of Computer Programs and Automated Data Systems for the Initiation Phase

NBS-FIPS-PUB 64, August 1, 1979, provides guidance in determining the content and extent of documentation for the initiation phase of the software life-cycle. It covers preparation of project request, flexibility study, and cost/benefit analysis documents.

4.2.3 <u>Flowchart Symbols and Their Usage in Information Processing</u>

This ANSI X3.5-1970, FIPS 24 and DOD standard establishes flowchart symbols and their usage in the preparation of flowcharts for information processing systems.

4.2.4 <u>Guidelines for Describing Information Interchange Formats</u>

NBS-FIPS-PUB-20, March 1, 1972, identifies and defines the physical and logical characteristics of formatted information to improve its interchange, processing, and use.

4.2.5 <u>Software Summary for Describing Computer Programs and Data Systems</u>

NBS-FIPS-PUB-30, June 30, 1979, establishes a standard form to be used by federal agencies in documenting summaries or abstracts of programs and automated data systems.

4.2.6 <u>Transmittal Form for Describing Computer Magnetic Tape File Properties</u>

NBS-FIPS-PUB-53, April 1, 1978, provides a standard form for federal agencies to use in documenting the physical properties and characteristics of a recorded magnetic tape file.

SECTION 5

5. OPERATING SYSTEMS

There are not, at present, any standards for computer operating systems, although there are several ongoing efforts, described below, on this subject.

It is difficult to assess, at this point, the relevance of the work on standardization of operating systems to DBMS. Although there seems to be an intention by the standardization committees to leave outside of their scope of work all data management functions, there can be other aspects of their work (like file naming conventions) that may be relevant to DBMSs. For this reason, the status of their work is described here.

5.1 NASA Transportable Applications Executive (TAE)

The Transportable Applications Executive (TAE) is a transportable software executive under development at the Goddard Space Flight Center. This executive will facilitate user access to data and programs used in remote sensing applications and will provide the necessary system control, bookkeeping, operating system interface, man-machine communications, and other services needed to simplify and standardize the user environment. A major design requirement of TAE is that it will have a high degree of "portability" (defined as a measure of the ease with which a program can be transferred from one environment to another). It will also be VICAR system-compatible.

The TAE will ultimately be used in a multicomputer network. Interprocessor communication capabilities are being incorporated to share data and resources.

The initial functional design of the TAE is presently being reviewed and refined. A first (partial) implementation is planned for completion by mid-1981 [61].

5.2 National Standards Organizations

ANSI committee X3H1 is currently developing command and response standards (job control language) for operating systems. The scope of work of X3H1 does not include any data management functions [62], but does include file naming conventions that can be relevant to database management systems (ANSI committee X3T5 - Open Systems Interconnection is also investigating file naming conventions).

It is expected that a proposal will be issued for public review by the end (September) of 1982 [63].

The NBS will also issue a proposal by the end of 1982.

The CODASYL group is expected to release a _CODASYL Common Operating System Command Language_ document by the end of 1980.

SECTION 6

6. DATA ADMINISTRATION AND AUDITING

There are no official standards published in this area and it is not expected that in the near future a proposal will be presented by any of the major standardization organizations.

The ANSI/X3/SPARC Study Group on Database Management identified, in its 1975 Interim Report [14] three roles in the management of data:

1) The enterprise administrator, with responsibility over the conceptual schema (see Section 2.2);

2) The application administrator, with responsibility over the external schema (see Section 2.2); and

3) The database administrator, with responsibility over the internal schema (see Section 2.2).

The same individual may function in different roles and one role may involve several individuals simultaneously. It is critical, however, that there is only one enterprise administrator and one database administrator, while there may be many application administrators.

Each administrator would be responsible for providing to the system a particular view of the necessary data, the relevant relationships among that data, and the rules and controls pertinent to its use.

SECTION 7

7. TERMINOLOGY

ANSI standard X3/TR-1-77 and FIPS 11-1 is a dictionary for information processing, in general, that was developed by combining existing lexicons, and also by studying the use of terms throughout the field of computers and information processing. The absence of database terms, however, is quite pronounced.

The International Standards Organization/Technical Committee 97 (ISO/TC97) is currently involved in developing a glossary for database management systems.

Task Group 24 (TG24) on Database Management Systems of the NBS has developed a terminology [64] for their internal use. This terminology definition, however, is not intended to be a standard.

SECTION 8

8. REFERENCES AND BIBLIOGRAPHY

[1] American National Standards Institute (ANSI). 1430 Broadway, New York, New York.

[2] American National Standards Covers the Wide World of Information Processing. Available from ANSI Reference [1].

[3] American National Standards Committee X3 - Computers and Information Processing. Computer and Business Equipment Manufacturers Association (CBEMA). Washington, D.C.

[4] Byers, R. A., "NASA Data Systems Standards". Proceedings of the International Conference on On-Board Data Management. Nice, France, October 24-27, 1978. European Space Agency, Paris, France.

[5] NASA Computer Resources Management Handbook. Report No. NMI 2410.1B National Aeronautics and Space Administration, Washington, D.C., March 1979.

[6] COBOL Extensions to Handle Databases. Report to the CODASYL COBOL Committee. Newsletter of ACM Special Interest Group in Business Data Processing. April 1978.

[7] CODASYL Database Task Group Report. Available from: ACM. April 1971.

[8] Proposal for a Database Facility in COBOL. CODASYL Database Language Task Group. Order from: Technical Services Branch, Department of Supply and Services, Ottawa, Canada, January 1973.

[9] CODASYL COBOL Journal of Development. Order from: Technical Services Branch, Department of Supply and Services, Ottawa, Canada, 1976.

[10] Olle, T. W., The CODASYL Approach to Database Management. John Wiley and Sons. 1977.

[11] "Implementation of the CODASYL DBTG Proposals". Proceedings of the British Computer Society Database Working Group. October 1973.

[12] June 1975 Report. British Computer Society/CODASYL Data Description Language Committee Database Administration Working Group.

[13] Lucking, J. R., Database Languages, in Particular DDL, Development at CODASYL. International Computers Limited, Systems Programming Division, Kidsgrove, Stoke-on-Trent, England.

[14] Interim Report. ANSI/X3/SPARC Study Group on Database Management Systems. February 1975.

[15] ANSI/X3/SPARC Database Systems Study Groups. Standing Paper 1. Version 1.1. T. B. Wilson, Editor. August 1980.

[16] "Report of the Study Group on Database Management Systems", The ANSI/X3/SPARC DBMS Framework. ANSI/X3/SPARC Database Systems Study Group. 1977.

[17] Recommendations for Database Management System Standards FIPS Task Group on Database Management System Standards. Special Publication 500-51. Institute for Computer Sciences and Technology. National Bureau of Standards.

[18] Jeffrey, S., Fife, D., Deutsch, D., and Sockut, G., Architectural Considerations for Federal Database Standards. Institute for Computer Sciences and Technology. National Bureau of Standards.

[19] Journal of Development. CODASYL Data Description Language Committee. January 1978.

[20] "The ANSI/SPARC DBMS Model". Proceedings of the 2nd SHARE Working Conference on Database Management Systems. Montreal, Canada, April 26-30, 1976 and North Holland 1977.

[21] "Draft Specification of a Data Storage Description Language". Appendix to the CODASYL Data Description Language Committee Journal of Development. British Computer Society/CODASYL Data Description Language Committee, Database Administration Working Group. January 1978.

[22] Journal of Development. CODASYL FORTAN Database Language Committee. FORTRAN Database Facilities. January 1980.

[23] Johnson, H. R., Standardization and the Relational Data Model. ANSI/X3/SPARC working paper RTG-HRJ-80-4. Relational Task Group. September 10, 1980.

[24] Data Dictionary Systems: Standardization? Database Newsletter, Vol. 8, No. 4, p. 2. Performance Development Corporation. July 1980.

[25] Prospectus for Data Dictionary System Standard. Document No. NBSIR 80-2115. National Bureau of Standards. United States Department of Commerce, Washington, D.C., September 1980.

[26] A Status Report on the Activities of the CODASYL End-User Facilities Committee (EUFC). ACM Signed Record, Vol. 10, Nos. 2-3, 1978. February 1979 (issued August 1979).

[27] desJardins, R. and White, G. W., "ISO/ANSI Reference Model of Open Systems Interconnection". Trends and Applications Symposium: Computer Network Protocols, IEEE-CS/NBS 1980. Gaithersburg, Maryland, May 29, 1980.

[28] McGovern, J. P. and Basu, D., "Middle Layers of Open Systems Interconnection: Session and Transport". Trends and Applications Symposium: Computer Network Protocols, IEEE-CS/NBS 1980. Gaithersburg, Maryland, May 29, 1980.

[29] Definition and General Issues of the VT Service. Draft. ANSI/X3T55/80-51, ISO/TC97/SC16/WG5/Berlin 44.

[30] Job Transfer and Manipulation Service. Draft. ANSI/X3T55/80-52, ISO/TC97/SC16/WG5/Berlin 45. (internal document.)

[31] Architectural Description of Upper Layers (Application and Presentation) of OSI Reference Model. Draft. ANSI/X3T55/80-53, ISO/TC97/SC16/WG5/Berlin 46.

[32] OSI - The Virtual File Service. Draft. ANSI/X3T55/80-54, ISO/TC97/SC16/WG5/Berlin 47.

[33] Common Command Language Feature Analysis. Draft, prepared by System Development Corporation. Institute for Computer Sciences and Technology. National Bureau of Standards. June 1980.

[34] Distributed Database Technology - An Interim Report of the CODASYL Systems Committee. A FIPS Conference Proceedings, Vol. 47. 1978.

[35] Commentary on the CODASYL Systems Committee Interim Report on Distributed Database Technology. A FIPS Conference Proceedings, Vol. 47. 1978.

[36] Swager, J., Architecture of a Distributed Information Resource. ANSI/X3/SPARC Database System Study Group. August 12, 1980.

[37] Truszkowski, W. and Posey, K., Database Interface With Distributed Systems: Global DDL, Global DML, Global Query. Draft - Version 1. July 6, 1979.

[38] Ferris, A. G. and Greene, E. P., A Proposed Concept for Improved NASA Mission Operations Data Management. Mission and Data Operations Directorate, X-533-76-81. Goddard Space Flight Center, Greenbelt, Maryland, October 1976.

[39] System Guidelines for Packetized Telemetry Formatting. NEEDS Program Office, Office of Space and Terrestrial Applications, National Aeronautics and Space Administration, Washington, D.C.

[40] Multimission Packet Telemetry Guidelines and Standards for Deep Space Missions. Issue 1. End-to-End Information System (EEIS) Program Office, Jet Propulsion Laboratory, Pasadena, California, November 8, 1979.

[41] Greenberg, E. and Hooke, A., Guidelines for Data Communications Standards. Standard Message Formatting for Space Flight Applications. Draft 3. Jet Propulsion Laboratory, Pasadena, California, July 6, 1981.

[42] Emmens, W. F., Network Control Protocols. IEEE COMPCON. 1978.

[43] Siedman, J. B. and Smith, A. Y., VICAR Image Processing System (Guide to System Use). Publication 77-37, Revision 1. Jet Propulsion Laboratory, Pasadena, California, December 1, 1979.

[44] Apenyo, K., Survey of Representative Database Systems. Draft. Jet Propulsion Laboratory, Pasadena, California, November, 1980.

[45] Martin, M., Proposed Multimission Imaging EDR Format. Interoffice Memorandum, SDP:384-80-318. Jet Propulsion Laboratory, Pasadena, California, March 11, 1980.

[46] User Computer Compatible Tape (CCT) Format Family Requirements. Document No. CCB-CCT-0001 A. Landsat Ground Station Operators Working Group (LGSOWG). October 16, 1978.

[47] The CCT Family of Tape Formats. Document No. CCB-CCT-0002 A. Landsat-D CCT Standards Committee. August 28, 1979.

[48] Landsat-D User Tape Format. Document No. FO-LSD-0001 A. Goddard Space Flight Center and EROS Data Center. August 28, 1979.

[49] Standard Format for the Transfer of Geocoded Information in Spatial Data Polygon Files. Document No. FOR-SDP-0001 A. Canada Centre for Remote Sensing. Energy, Mines and Resources, Canada, July 1979.

[50] Proposed American National Standard Specification for an Information Interchange Date Descriptive File. Draft. ANSI BSR X3.87. March 24, 1980.

[51] Brooks, A. S., Programme 97.15.6: Interchangeable IRV Coded Data Files Technical Background. Document No. 97/15/3 N29, 97/15 N79. International Standards Organization. July 12, 1979.

[52] Workshop on Standards for Image Pattern Recognition. Proceedings held at the National Bureau of Standards, Gaithersburg, Maryland, June 3-4, 1976. National Bureau of Standards. 1977.

[53] Dehue, J. S., "The NATO RSG-4/SGIP Tape Format". pp. 11-23. Reference [52].

[54] Pavlidis, T., "Tape Formats". pp. 26-27. Reference [52].

[55] Initial Graphics Exchange Specification. IGES Version 1.0. Programmable Automation Section, National Engineering Laboratory. National Bureau of Standards, January 1980.

[56] NASA Software Management Guideline. Appendix G of Reference [5].

[57] Durachka, R., Banks, M., Witomen, K., Posey, K., and Cyprych, G., "Software Engineering Standards and Software Development Life Cycle". Information Extraction Division, Goddard Space Flight Center, Greenbelt, Maryland, January 1980. (internal document).

[58] Implementation and Management of JPL Software. Document No. 500-152. Jet Propulsion Laboratory, Pasadena, California, June 25, 1980. (JPL internal document).

[59] JPL Software Standards Study and Follow-on Activity Plan. Document No. 500-175, Jet Propulsion Laboratory, Pasadena, California, July 7, 1980. (JPL internal document).

[60] Guideline for Planning and Management of Database Applications. FIPS PUB 77. National Bureau of Standards. September 1, 1980.

[61] Van Wie, P., Fisdel, D., and Howell, D., _A Transportable Executive System for Use With Remote Sensing Applications Software_. Goddard Space Flight Center, Greenbelt, Maryland. Presented at the 1980 Machine Processing of Remotely Sensed Data Symposium, March 1980.

[62] Sockut, G. H. and Brown, J., Jet Propulsion Laboratory, Pasadena, California. (private communication).

[63] _"A Family of Standards for Database Management"_. National Bureau of Standards, Gaithersburg, Maryland, September 8-10 1980.

[64] Appendix 3 of Reference [17].