NBS TECHNICAL NOTE 887

U.S. DEPARTMENT OF COMMERCE / National Bureau of Standards

Six Data Base Management Systems: Feature Analysis and User Experiences

NATIONAL BUREAU OF STANDARDS

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

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

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

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

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

THE INSTITUTE FOR APPLIED TECHNOLOGY provides technical services to promote the use of available technology and to facilitate technological innovation in industry and Government; cooperates with public and private organizations leading to the development of technological standards (including mandatory safety standards), codes and methods of test; and provides technical advice and services to Government agencies upon request. The Institute consists of the following divisions and Centers:

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

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

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

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

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

---

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

[2] Located at Boulder, Colorado 80302.

Six Data Base Management Systems: Feature Analysis and User Experiences

Elizabeth Fong, Joseph Collica
and Beatrice Marron

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

Sponsored in part by:

Federal Aviation Administration
Office of Management Systems
800 Independence Avenue, S.W.
Washington, D.C. 20591

U.S. DEPARTMENT OF COMMERCE, Rogers C. B. Morton, *Secretary*

James A. Baker, III, *Under Secretary*

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

NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Acting Director*

Issued November 1975

National Bureau of Standards Technical Note 887

Nat. Bur. Stand. (U.S.), Tech. Note 887, 84 pages (Nov. 1975)

CODEN: NBTNAE

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1975

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402
(Order by SD Catalog No. C13.46:887). Price $1.45 (Add 25 percent additional for other than U.S. mailing).

PREFACE

The principal motivation for this report was a request from the Federal Aviation Administration for NBS assistance in assessing several data base management systems for their relative applicability and suitability for meeting FAA's Accounting System requirements.

We gratefully acknowledge the excellent cooperation of the systems vendors and Federal users, both in providing information and validating the technical content of this report. We especially appreciate the support of Dr. Dennis Fife, Chief, Computer Science Section, NBS, and Mr. Robert Link of the Office of Management Systems, FAA.

This report identifies the data base management systems by trade names as necessary to provide a descriptive characterization of their features. Inclusion of a system in this report in no case implies a recommendation or endorsement by the National Bureau of Standards, and the presentation should not be construed as a certification that any system provides the indicated capabilities. Similarly the omission of a system does not imply that its capabilities are less than those of the included systems. The information presented was obtained primarily from vendor's documents and has been reviewed by each vendor for accuracy and clarity; the authors retained final technical judgement on the information included. The report is intended to be informative and instructive in state of the art assessments, and not an evaluation of the systems.

Due to the dynamic nature of the systems' features and the user application environment, the information presented is current only to April 1975.

TABLE OF CONTENTS

SIX DATA BASE MANAGEMENT SYSTEMS: FEATURE ANALYSIS & USER EXPERIENCES

Elizabeth Fong, Joseph Collica, and Beatrice Marron

This report presents an objective overview of features of six selected data base management systems (DBMS) and Federal user experiences with these systems. Application criteria were developed in order to aid in the evaluation and selection of DBMS.

The advantages of generalized DBMS over traditional methods of software system development are discussed. The criteria for choice of the six systems' features are presented. The data reported were gathered from two sources: vendors and users. Vendor information consisted of side-by-side presentation of features of the six DBMS. User experiences reported are summarized under appropriate headings. This information is used to derive application criteria for assessing the usability and operational suitability of DBMS to a variety of data processing requirements.

The six systems reviewed are: ADABAS, IMS/VS, INQUIRE, MODEL 204, SYSTEM 2000, and TOTAL.

Keywords: Data base applications; data base management systems; selection criteria; software features; system evaluation; system features; user experiences.

1.0 INTRODUCTION

1.1 Background

In the last ten years, the commercial availability of generalized data base management systems (DBMS) has increased markedly. Despite a history marked by some significant failures and numerous successes, the field of DBMS for data processing applications now offers a wide and growing range of choices. Compounded with this proliferation is the continuous enhancement and tuning of a package by its vendor to make it as competitive as possible. Thus, keeping step with the latest development and state of the art of DBMS is difficult. Yet, in the evaluation and selection of these DBMS, accurate and up-to-date technical details need to be studied. This report offers relief as a reference to technical features and operational experiences in the domain of generalized DBMS.

The principal motivation for the NBS study was a request for assistance from the Federal Aviation Administration* in assessing

*Federal Aviation Administration Interagency Agreement #DOT-FA74WAI474

several data base management systems for their relative suitability in meeting FAA's accounting system requirements. However, the study has more general applicability in that it offers application criteria, as well as feature analysis and user experiences in the evaluation and selection of DBMS.

1.2 Advantages of DBMS

For the purpose of this study, DBMS are characterized as generalized software, which provides a single flexible facility for accommodating different data files and operations while demanding less programming effort than conventional programming languages. DBMS possess the following general properties:

- Software which facilitates operation on data such as data definition, data storage, data maintenance, data retrieval, and output.
- Software which facilitates reference to data by name and not by physical location.
- A software environment which is not tied to a particular set of application programs or files.

Systems vary considerably with respect to their features. There are the more elementary systems such as sequential file systems having single record structures and providing only rudimentary report formatting facilities. In contrast, there are very elaborate systems handling several files, performing functions in an on-line mode, and having sophisticated query and report writing capabilities.

In February 1974 Computerworld [1]*reported the growth of DBMS as follows:

> "In 1970, there were perhaps 100 users of DBMS in the U.S. Today, that number is anywhere from 800 to over 1,000."

Figure 1 shows a dramatic 53% increase during six months in 1974 in the number of users of the six DBMS included in this report. Thus, the use of DBMS to control large data bases and provide information to multiple users has already gained acceptance in the data processing world. This is partly due to the growing body of massive data collected, for example, there are over 450 distinct data files that are available for public sale through National Technical Information Service. Another reason is the recent hardware advances in storage devices, randomly accessible disks, laser and bubble memories, etc., which make storage of large amounts of data feasible.

*Figures in brackets indicate the literature reference at the end of this paper.

Name of DBMS	No. of Users as of Feb. 1974	No. of Users as of Oct. 1974
ADABAS	25*	30**
IMS	400*	500**
INQUIRE	40***	49***
MODEL 204	6***	21***
SYSTEM 2000	60*	70**
TOTAL	400*	750**

* Figure taken from reference [1]

** Figure taken from reference [2]

*** Figure quoted from vendor

FIGURE 1 - Number of Users of Six DBMS

Historically, the method of establishing and processing large amounts of data has been sequential and tape or card oriented. Specialized computer programs were developed primarily in COBOL to store and retrieve the data. In the last decade, as massive quantitites of data have proliferated, development has turned to generalized software that is independent of any particular file formats, allowing users to define and process a variety of distinct applications.

In the evaluation and selection of a software facility for performing data processing functions, one can immediately think of two extreme alternatives:

(1) Design and implement the system by tailoring it for a specific application without using any pre-packaged software. For example, one could write a data base management system in assembly language or in COBOL.

(2) Utilize a generalized off-the-shelf packaged DBMS and build any necessary additional functions with application programs.

The first way suggests an initial investment in the data base system design and implementation. There is a long time lag before the system is completed for use. Another drawback is that this system is tailored to a particular application, and may not be flexible and responsive to external changing requirements.

There are many advantages in taking the second alternative. These are:

- A shorter period of time is required to develop a working system.

- Less manpower is required to install and support an off-the-shelf package than writing software from scratch.

- The personnel resources can be better utilized in doing problem-oriented analysis rather than detailed coding of the system.

With the packaged DBMS approach, there are various levels of supporting effort in the software implementation task.

In general, generalized DBMS software can be regarded as a layer of software which performs user data services. The functions which a modern DBMS provides can be identified as follows:

- Accepts a data definition from the user. This is usually referred to as logical data.

- Maps logical data onto the physical storage devices by developing a data organization scheme.

- Manages the storage space and the data base.

- Performs data manipulation functions such as retrieval of data for display, retrieval of data for reports, changing data, adding data, and deleting data.

- Provides languages for human interface which usually are either English-like or host language calls (COBOL, FORTRAN, or PL/1) with appropriate linkage to the DBMS.

- Provides various controls and checks for data validity, integrity, security and privacy, plus various utility programs.

Figure 2 depicts the functional components of a generalized DBMS. A user defines the data to a DBMS via the provided data definition language. From the raw data and the data definition, the DBMS creates the data base on a physical mass storage device. The data management functions are accomplished by software which translates or interprets user requests for delivering retrieval records or for performing the necessary maintenance activities for the data base.

All the off-the-shelf DBMS packages do not provide the same set of functions, and the implementation of functions differs widely in depth and strength of effectiveness. A careful evaluation and selection effort is crucial. But without all the facts about each package, an agency has great difficulty in making a choice. The following chapters present facts about the features of six DBMS, and a description of how

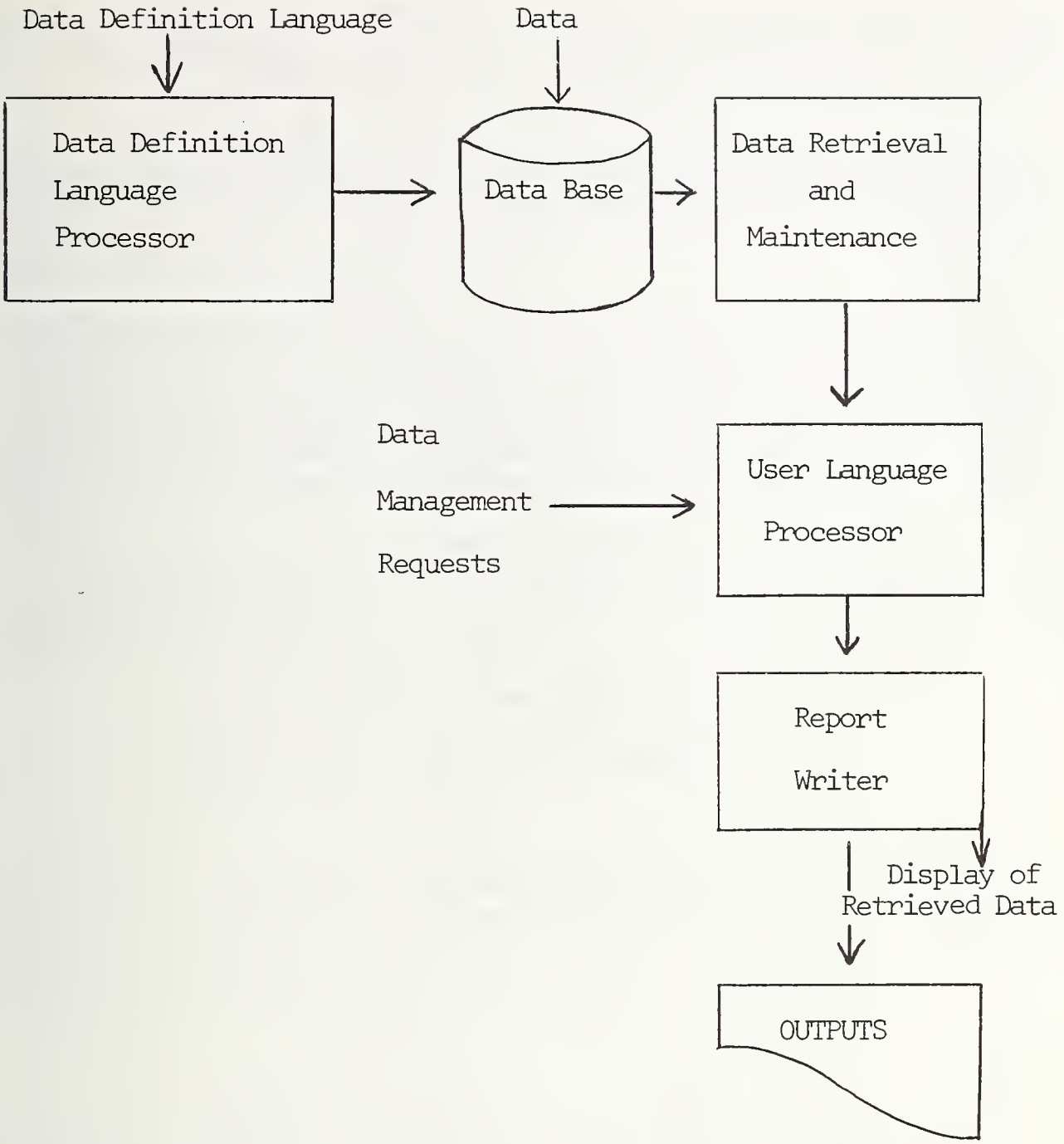

FIGURE 2. DBMS Functional Components.

they are being used in real application environments. Due to the dynamic nature of the systems' features and the user application environment, the information presented is current only to April 1975.

There is no simple answer to which DBMS is the best choice because the advantages and disadvantages need to be considered in light of an installation's ultimate needs.

1.3 Approach

Six modern DBMS were chosen for this study. Features of all six systems are contrasted side-by-side in Section 2. The data here on each of the systems were obtained primarily from system documentation. In cases where the available documentation did not provide sufficient data, further information was obtained by telephone from the vendor. The technical information on each system was validated by the vendor.

Feature analysis of systems, however, covers only half of the picture--the vendor's side. In order to assess the effectiveness of a system, the user's side must also be considered. The data gathered from users include answers to the following management-type questions:

- Why the specific DBMS package was chosen?
- How easy (or difficult) was installation?
- How fast did the staff learn to use it?
- What level of staff expertise is required to operate and use the package?
- On what computer configuration is the DBMS installed?
- How many application programs are necessary?
- How good is the documentation and training provided?
- How responsible is the vendor - in installation, in full implementation support, in fixing bugs, in system enhancement?

Rather than taking a broad sample of the user community for each system via qustionnaires, it was decided to limit the user study to in-depth personal interviews with two Federal agency users of each system*. Each of the interviews is reported in Section 3 with uniform headings. The information from each installation was reviewed and

---

*There is only one government user of the ADABAS system at this time.

validated by the agency. The user experiences and overall impression of these interview notes reflect some subjective opinion by the agency as interpreted by NBS.

The system feature analyses and the user experience information were then applied to derive application criteria for the systems included in this report. These application criteria are presented in Section 4 and provide a guide for the selection and evaluation of DBMS based on application considerations.

1.4 Choice of Six DBMS

There are many surveys of DBMS packages (See Figure 3). The number of systems covered ranges from six to as many as 154. Unforunately, because of the extremely dynamic and rapidly evolving character of the field, most of the surveys that were conducted before 1974 are essentially outdated.

The criteria for inclusion in the NBS study limit the scope to those systems that are:

- Capable of processing formatted data, that is, data elements are mostly numerical, abbreviated text, or encoded values.
- Systems that are actively marketed.
- Systems that have more than 20 users.

Other criteria included the additional requirements imposed by FAA limiting the number of systems. Six packages were selected for this study: ADABAS, IMS/VS, INQUIRE, MODEL 204, SYSTEM 2000 and TOTAL. These six systems are believed to be representative of widely recognized and proven products available in the market today.

2.0 FEATURE ANALYSIS OF DBMS

In this section the features of the six DBMS are identified and contrasted along the same set of characteristics. The feature list approach for the common description of DBMS is patterned after the CODASYL Technical Report of May 1971 [4]. The six systems are ADABAS, IMS/VS*, INQUIRE, MODEL 204, SYSTEM 2000, and TOTAL.

The feature descriptions provided below are presented under nine headings: general description, computer environment, data structure, data definition, data maintenance, query specification, output and report generation, security features and external linkages. Each

---

*In the case of IMS, several versions existed: IMS/II, IMS/VS, DL/1 DOS/VS, with IQF (Interactive Query Facility) and GIS (Generalized Information System). IMS/VS with GIS was chosen here.

		No. of Systems
CODASYL		
1969 -	A Survey of Generalized DBMS [3]	9
1971 -	Feature Analysis of Generalized DBMS [4]	9
NBS		
1972 -	Index of DMS Packages [5]	154
1971 -	A Survey of Selected Document Processing Systems [6]	8
1974 -	A Technical Index of Interactive Information Systems [7]	46
1973 -	A Study of Six University-Based Information Systems [8]	6
MITRE		
1969 -	DMS Survey (MIP 329) [9]	10
1973 -	DMS Catalog (MIP 139) [10]	61
STANFORD		
1974 -	A Feature Analysis of Interactive Retrieval Systems [11]	11
DATAPRO		
1974 -	A Buyer's Guide to Data Base Management System [2]	6

FIGURE 3. Surveys of DBMS Packages

heading is further subdivided as necessary to describe the capabilities and functions appropriate to each heading. This is followed by a table covering the capabilities of each of the six systems. Every effort was made to insure correctness of the data; the feature comparison was prepared from documentation obtained in user manuals and contacts with vendor representatives. It was not based on extensive hands-on experience.

1. General Description - This section provides the proper identification and background information about each system. Availability, cost, and type of applications are all covered.

2. Computer Environment - The hardware and software configurations are covered in this section. Systems are built to interface into a particular operating system, a particular teleprocessing communications interface, and mass storage with input/output access method package. An important user consideration is the mode or modes of operation available for each system.

3. Data Structure - Generalized data base management systems can handle at least two levels of data structures: the logical data structures as seen by the user of the system, and the physical data structure which is the mapping of the data into storage. Each system also supports a set of legal data types upon which legal operations can be defined.

4. Data Definition - The definition of each level of data in the data structure is covered in this section. Systems differ in terms of the amount of user control.

5. Data Maintenance - The creation of the data base, the restructuring process and various aspects of updates are covered in this section. Data creation provides the initial instance of the data base. It is the process of reserving storage, establishing item names and relationships as defined in the data definition language. Restructuring is the changing of data definition. Update is the process of changing the value content of some part of the data base without changing the structures.

6. Query Specification - Query specification is a process of selecting and extracting some parts of the data base for display. There are many degrees of sophistication in terms of query formulation and various degrees of user interaction are required.

7. Output and Report Generation - The output is usually the display of some parts of the data base as a result of query. The display ranges from system-provided output to user formatting of reports. It is during this stage that special operations such as sorting and various kinds of computations may be applied on the selected data as a built-in process algorithm.

8. Security Features - Two aspects of security are addressed: protection of systems against hardware failure, loss of data bases, etc., and the protection of systems against unauthorized users. The former is usually considered in terms of the procedure for backup, with restart recovery mechanisms. The audit trail mechanisms not only serve as a means of monitoring the daily transactions but can be used for data base recovery. The protection against the unauthorized user is usually via password keys with various types of authorization assigned to those keys.

9. External Linkages - To be of general use, it is necessary to have a facility whereby the generalized data base management system functions can be accessed via a procedure language application program such as COBOL, FORTRAN, or PL/1. The variety of the external languages provided is described under the procedural language interface section.

2.1 General Description

2.1.1 System Name, System Originator, Address, Telephone

ADABAS	Adaptable Data Base System Software Ag 11800 Sunrise Valley Drive Reston, Virginia 22091 (703) 620-9577
IMS/VS	Information Management System/Virtual System International Business Machines Corp. Data Processing Division 112 East Post Road White Plains, New York 10601 (914) 949-1900
INQUIRE	INQUIRE System Infodata Systems, Inc. 5205 Leesburg Pike Falls Church, Virginia 22041 (703) 578-3430
MODEL 204	MODEL 204 Computer Corporation of America 575 Technology Square Cambridge, Massachusetts 02139 (617) 491-3670
S2000	System 2000 MRI Systems Corp. P. O. Box 9968 Austin, Texas 78760 (512) 258-5171
TOTAL	TOTAL System CINCOM Systems, Inc. 2300 Montana Avenue Cincinnati, Ohio 45211 (513) 662-2300

2.1.2 Availability and Cost of Software. Different software distribution plans exist. Software to be installed at a customer's site may be available by purchase or lease. Software may also be available from time-sharing service centers. The rental and purchase costs for each of these systems were obtained from the GSA Schedule where available and from the software vendors. In certain cases, a discount is allowed if more than one copy of the system is acquired within the same government agency.

	Availability	Cost
ADABAS	Software available for purchase from Software Ag Lease arrangement from Software Ag Time sharing service available from Planning Research Corp.	$90,000 under GSA Schedule Six months to five year lease available. Training $2,000. $3,375/month under GSA Schedule.
IMS/VS	Software available for lease only from IBM	IMS/VS $770/month IMS/VS DC feature $935/month IMS-2 $616/month (OS) IMS-2 DC $700/month GIS/VS $800-1,500/month
INQUIRE	Software available for purchase from INFODATA Available on a rental basis from INFODATA. Time-sharing service available from Boeing Corp.	$32,000 for basic INQUIRE--under GSA Schedule. Numerous options available. $1,325 per month for 29 months under GSA Schedule.

MODEL 204	Software available for purchase from Computer Corporation of America	$25,600 for nucleus system Additional modules increase cost to $82,500 for full system
	Lease arrangement from Computer Corporation of America	$2,600/month for full system
	Time-sharing service available from Informatics, Inc., Fairfield, N.J.	
S2000	Software available for permanent paid up lease from MRI	$27,000 Basic System under GSA Schedule. Numerous options available.
	Various lease arrangements from MRI (rental with option to buy)	$1,140/month Basic System under GSA Schedule.
	Time-sharing service available on INFONET, CDC, UCS, COMNET, TYMSHARE, Planning Research Corp., Boeing Computer Services, and others.	Same as regular GSA INFONET charges except no discount and $.03/512 wds/day
TOTAL	Software available for purchase (25 year paid-up lease).	Purchase price $32,495 under GSA Schedule.
	Software available for annual lease.	$1,050/month minimum package under GSA Schedule. There are many different TOTAL versions.

2.1.3 History of Software. Historical information on the implementation of the software indicates extent of development and field usage. The specific information concerns the date of first installation and the name of significant installations.

	History and Dates	Significant Installations
ADABAS	Software development started by Software Ag of Darnstadt, W. Germany in 1968. The first commercial installation in March 1971. Marketed in U.S. by Software Ag North America since early 1972.	First National City Bank of N.Y., Shell Oil, International Nickel Corp.
IMS/VS	Software development occurred as a joint venture between IBM and North American Rockwell Co. in mid-1960's. IMS Ver 1 1969, IMS Ver 2 1971, IMS/VS 1973	Federal Reserve Board U.S. Navy
INQUIRE	The system was developed primarily for pharmaceutical environment. The first installation was around 1969	National Library of Medicine, LEAA, DOT, CIBA-GEIGY Corp., Exxon
MODEL 204	The software is a proprietary development of CCA. The first installation was in 1971.	Food and Drug Admin. Parklawn Computer Center State Department Department of Commerce
S2000	System is based, in part, on work performed at the University of Texas at Austin on Remote File Management Systems (RFMS). System first installed in July 1970.	EPA, Federal Power Commission, HUD, U.S. Geological Survey, Chrysler, Ford Motor Co.
TOTAL	TOTAL was designed in early 1968	Social Security Admin., Bureau of Labor Statistics, Exxon, Ford Motor Co.

2.1.4 Type of System and Application. A common characterization of system types is self-contained versus host-language. Self-contained data management systems typically provide all essential capabilities for data creation, data update, data retrieval and report formatting. These self-contained systems usually provide a generalized retrieval language that can express fairly complex selection criteria. On the other hand, the host-language data management systems are characterized by embedding system functions within a host-language program such as COBOL, FORTRAN, PL/1 or assembly language. The user must then rely on the host-language to program some of the functions.

Even though most of the data management systems are generalized, meaning there exists a data definition facility whereby a user can define his own application data, certain features make a system particularly suited for structured text searching such as in bibliographic information or legal applications. Some systems were designed to handle formatted data where data elements have mostly numeric, abbreviated text, or other encoded values. Examples of such types are inventory control application, personnel application, etc.

	Type	Application
ADABAS	Self-contained system	Formatted data processing
IMS/VS	IMS/VS is a host-language system. GIS is a self-contained system. IMS with GIS is both a host-language and self-contained system.	" " "
INQUIRE	Self-contained system	" " "
MODEL 204	Self-contained system	" " "
S2000	Self-contained system	" " "
TOTAL	Host-language system	" " "

2.2 Computer Environment

2.2.1 Main Frame, Operating System, Core Requirements and Source Language. The main frame refers to the particular computing machine for which the system was designed. The name of the operating system which supports the DBMS and the main frame is then listed. The core memory requirements may vary depending on the number of users, and the amount of input-output buffer space in the system. The amount cited is the minimum core required. The source language is the programming language in which the system itself is written.

	Main Frame	Operating System	Minimum Core Requirements	Source Language
ADABAS	IBM 360/370 Series	OS or DOS OS/VS1 OS/VS2	110K bytes with no overlays	IBM 360/370 assembly language
IMS/VS	IBM 370 Series	OS (IMS-2) OS/VS (IMS/VS)	512K bytes data base and data communications for OS 786K bytes for data base and data communi-cations under VS	IBM 360/370 assembly language
INQUIRE	IBM 360/370 Series	OS, VS 1, VS 2	120K to 200K bytes	PL/1 and assembly language
MODEL 204	IBM 360/370 Series	OS or OS/VS	120K bytes of core 225K bytes with average buffer space 300K for largest system	IBM 360/370 assembly language

S2000	Univac 1100 Series	Exec 8, CSTS	31K decimal words	Primarily in assembly language plus high level FORTRAN. On IBM, Assembly language plus MRI's own system programming language
	CDC 6000 Series	SCOPE, KRONOS	20K words	
	IBM 360/370 Series	OS, OS/VS	160K-200K bytes	

TOTAL	IBM 360/370 Series	DOS, DOS/VS, OS, OS VS1/VS2	4-12K bytes plus buffers	Assembly language
	Honeywell 200, 2000	MSR Mod 1 052000	4-12K bytes plus buffers	
	Univac Series 70 9400, 9700 Series	054, 059	4-12K bytes plus buffers	
	NCR Century Series	B1, B2, B3, B4	15K plus buffers	
	CDC 6000	SCOPE, KRONOS	15K plus buffers	
	IBM System/3 Model 10 Model 15	CCP CCP	9K plus buffers	

2.2.2 Communications. Communication establishes the link between the remote terminal user and the system. For a particular type of terminal, there is usually a teleprocessing interface program which processes a user's request and properly transmits the messages to and from the computer system. If the system is re-entrant (code does not modify itself), then multiple users can simultaneously be executing the one copy of the program. The data communication system may also operate in single-thread transaction or multi-thread mode. Single-thread transaction is the execution of application programs until all requirements of the transaction have been satisfied and the end result delivered to the user via the communication system. A multi-thread capability allows for concurrent execution of application programs.

	Terminal Type	Teleprocessing Interface Program	Re-entrant for Multiple Users	Teleprocessing Mode
ADABAS	Any compatible with TP software	Does not offer a TP system; INTERCOMM, TASK/MASTER, CICS, ENVIRON/1, IMS/DC and TSO have been set up to interface with it	Yes, except nucleus module	Multi-thread
IMS/VS	IBM 1050, 2260 local/remote, 2265, 2740-I or II, 2741, 2770, 2780, 2980, 3270 local/remote, 3600, 3790, system/3, system/7, TWX, SYSIN, SYSOUT	DC, the data communications feature, a module which can be ordered for use with IMS	YES	Multi-thread
INQUIRE	Any compatible terminal with TP software	It has its own TP. Also capable of interfacing into TSO, TCAM, and IMS/DC	Partially re-entrant	Single-Thread

MODEL 204	TTY up to 300 baud IBM 2741 or DATEL IBM 2265 & 2260 - remote	INTERCOMM, CICS, TSO, WYLBUR/MILTEN, TCAM.	Yes, modular re-entrant	Multi-thread
S2000	TTY 33, 35, IBM 2741, IBM 2260 CRT	TP2000 (optional) teleprocessing monitor. Other teleprocessing monitors such as ALPHA, HYPER-FASTER, BEST, INTERCOMM, and CICS may be used as interface at additional cost	The IBM version is re-entrant	Multi-thread
TOTAL	All IBM compatibles plus 3270 local/remote and many others	ENVIRON/1, CICS INTERCOMM, TASK/MASTER	Yes	Multi-thread

2.2.3 Mass Storage and I/O Access. Data Management software usually relies on other system software support for handling large volumes of data. This support includes mass storage devices (where the data base resides) plus the input-output access method packages. These packages are provided on the operating system of the machine or are separately available.

	Mass Storage Device	I/O Access Method Package
ADABAS	All removable disk packs are supported but no tapes	BDAM
IMS/VS	2311 disk 2314 " 2319 " 2305 " 3330 " Magnetic tapes	VSAM, ISAM OSAM, BSAM SAM
INQUIRE	3330, 2314, 2311 type disks 2301 drum or 2305 fixed head storage or equivalent devices supported by OS	BDAM and ISAM
MODEL 204	3330 disk 2314 " 2311 " Magnetic tapes	EXCP QSAM for sequenced I/O
S2000	Disk and magnetic tapes	BDAM & BSAM (IBM version)
TOTAL	2311, 2314, 3330, 3340, NCR disks, Honeywell disks, CDC disks	BDAM, DAM

2.2.4 Mode of Use. DBMS are designed to be operated in batch mode as well as on-line with interactive capabilities. The available mode (batch or on-line) might be administratively restricted at a given time in order to use system resources efficiently.

	Mode for Data Definition	Mode for Data Maintenance	Mode for Querying
ADABAS	Batch	On-line or batch	On-line or batch
IMS/VS with GIS	On-line or batch	On-line or batch	On-line or batch
INQUIRE	On-line or batch	On-line or batch	On-line or batch
MODEL 204	On-line or batch	On-line or batch	On-line or batch
S2000	On-line or batch	On-line or batch with immediate access feature; batch with host-language	On-line or batch with immediate access feature; batch with host-language
TOTAL	Batch	On-line or batch	On-line or batch

2.3 Data Structures

2.3.1 Data Types. Within each system there is a set of valid data types which can be defined for any element of a file. Examples are: alpha, alpha-numeric, numeric, date or coded (different internal and external representations for a given data value), etc. The definition of types prescribes certain legal operations upon the data values for a particular element or item.

	Data Types
ADABAS	Alphanumeric, alpha, binary, fixed point, floating point, unpack decimal, pack decimal
IMS/VS	All data types supported by the hardware alphanumeric, packed decimal, binary
INQUIRE	Alphanumeric, alpha, numeric, packed decimal, unpacked decimal, coded data and binary
MODEL 204	Alphanumeric, numeric (decimal integers, decimal fractions or both) text (up to 255 characters per field)
S2000	Names (extraneous blanks removed), text (retains blanks), date in form of all permutations of mm/dd/yy), integer, decimal, and money. All numeric fields stored as packed decimal.
TOTAL	All data types supported by the hardware: alphanumeric, packed decimal, binary.

2.3.2 Logical Data Structures. Logical data structures are the composition of the data items in a record as it is viewed by the user without regard to the way data are stored within the computer. Data may be defined to be hierarchical or tree-like allowing subelements, repeating group allowing subelements to occur zero or any number of times, and variable length text field. Some systems permit network relationships or free associations of data rather than strict hierarchical relationships. There is also the ability to interrelate the data across file and among records by defining associations among fields in the records.

	Logical Data Structures
ADABAS	One level file structure with repeating groups. Variable length text fields allowed with a limit of 253 bytes per field. One can define multiple fields up to 191 fields. Multi-level hierarchical network relationships can be achieved by coupling two files together.
IMS/VS	The logical data structure of IMS is based upon segments (repeating groups). An IMS data base consists of 1 to n records, a record consists of 1 to n segments. Hierarchical structures with up to 15 segment levels and 255 segment types per logical data base record. Interrelated tree structures allow a segment to participate in more than one logical data structure. Repeating groups and variable length text allowed.
INQUIRE	With one data base, there is a hierarchy of one level (parent-child). However, a multi-data base can be defined to have up to 32 levels of hierarchical structure. Repeating groups and variable length text (up to 32,760 bytes) are allowed.
MODEL 204	One level file structure with repeating groups variable length text fields allowed with a limit of 255 characters. Records are related by field values, and the relationships may represent trees, networks, rings, etc.
S2000	Hierarchical structures of records (called repeating groups) variable length alphanumeric fields allowed. The maximum number of levels is 32 hierarchical levels with no restriction on the number of repeating groups but the total components per data base may not exceed 1000. The maximum length of the alphanumeric field is 250 characters. There is no limit to the number of data bases. Data bases may be interrelated, providing full network structuring.
TOTAL	Two types of records are permitted: a single-entry or master record and a variable-entry record. 2,500 different access points into a variable entry record and 2,500 different record types within a variable entry file are also allowed. Inverted list, hierarchical structures and network relationship are also supported. Variable length text fields are supported logically.

2.3.3 Physical Storage Structures. The individual records are generally organized into some form of storage structure defined as a data space. Within this data space records are distributed in accordance with a defined algorithm. These algorithms dictate the search strategies for each system. The techniques for the organization of data values may be:

- Sequential where data records are arranged in a serial manner on the mass storage device.

- Indexed Sequential where data records are partitioned into smaller groups. Each group location is identified by an index, and records in a particular group are sequentially arranged.

- Inverted Structure and Chain List refer to the way keys (searchable data elements) are maintained. They are like indexed sequential except the index is the keyed data element. More than one index may be established.

- Random Organization usually refers to the technique of hash-coding the record value in order to determine the location of the record.

	Physical Storage Structures
ADABAS	All keyed fields are inverted and a search of the list will produce an internal storage reference. If non-key fields are specified, a message will occur.
IMS/VS	Two basic physical storage organizations are supported: Hierarchical Sequential, available through HSAM and HISAM; and Hierarchical Direct, available through HDAM and HIDAM.
INQUIRE	Physically three files are maintained: the data file containing user specified data, the search file containing chained subsets of data file together, and the index file containing keyed or inverted data elements. Keyed retrieval goes through the threaded list of search files which points to the item in the data file. Non-keyed retrieval scans the data file sequentially. Keyed and non-keyed data may be intermixed and used for retrieval.
MODEL 204	All keyed fields are inverted and a search of the list will produce an internal storage reference. Non-keyed fields cannot be used in retrieval, but can be tested after a record has been retrieved.
S2000	All keyed fields are inverted and searching the list will produce pointers to the hierarchical structure table to resolve multi-level data base searches. Non-keyed fields can be searched by entering the "Queued Access Module" and the system performs a sequential search after first isolating a subset of the data base through any keyed field specified.
TOTAL	Random organization in the main file and bidirectional hierarchical structure for the variable entry files which are accessed from the main file.

2.4 Data Definition

2.4.1 User Definable Fields. Any generalized data management system provides the user with a data definition language. The language form may be highly tabular, card-column oriented, or free form. Data naming, types, characteristics, and relationships with other data items are usually indicated at this time. For most systems, key and non-key fields usually must be indicated. Security restrictions may be indicated at definition time.

	User-definable Data definition	User-definable Keyed Fields	User-definable Security Restriction	Data Definition Language Form
ADABAS	Yes	Yes	Yes, by invoking a utility program call	One card per field description
IMS/VS	Yes	Yes	Yes	Free form
INQUIRE	Yes	Yes	Yes	Form is provided
MODEL 204	Yes	Yes	Yes	Attribute and value(s) specification
S2000	Yes	Yes	Yes	Syntax-directed (free form with punctuation symbols as delimiters or qualifiers)
TOTAL	Yes	No	No	List-like statements

2.5 Data Maintenance

2.5.1 Data Base Loading and Input Analysis. The data base management system usually takes the data definition and builds internal tables. At data base loading time, a certain amount of input analysis is provided. The input validation may be merely matching of data types or preprogrammed range checks and built-in data conversion programs. For certain text-oriented systems there are some facilities for vocabulary building, and automatic or machine aided context indexing.

	Input Media for Loading	Validation of Data Types	Preprogrammed Range Checks	Data Conversions
ADABAS	Sequentially organized records such as tape or cards	Yes	Yes	Yes, including encryption
IMS/VS	Tape, disc or cards	Not automatically accomplished	Not automatically accomplished	Not automatically accomplished
INQUIRE	Card, tape, disc, or any sequential organization	Yes	Not automatically accomplished	Not automatically accomplished
MODEL 204	Tape, disc or cards	Yes	Yes, use stored user language requests	Yes, translations occur via translation table
S2000	Sequentially organized records on tape/disc, or data preformatted into "value-string" form	Yes	Yes, if procedural language program is used	Not automatically accomplished
TOTAL	Cards, tape, disc	Not automatically accomplished	Not automatically accomplished	Not automatically accomplished

2.5.2 Data Base Restructuring Process. Restructuring process is the modification of data definition, such as adding/deleting a new data element, adding/deleting another level of hierarchy, adding/deleting a repeating group field, etc. One way of doing this is to unload the data, re-define and load again. Utility programs are sometimes furnished by the system. However, certain systems allow limited restructuring without an external program. A fully automatic restructuring process is the capability to modify existing data structure via a language.

	Restructuring Process
ADABAS	Total automatic restructuring allowed. User may add or delete single data element via a command, the system will add or delete the element from the data definition table. Values for each data are added or deleted using normal update commands.
IMS/VS	Restructuring allowed via system utilities and must be invoked by JCL.
INQUIRE	Only field expansions are allowed. Changing structure of field requires unloading and reloading.
MODEL 204	User may add or delete data elements, or records. Once assigned, field types cannot be changed without reloading the file.
S2000	Field types can be changed, repeating groups can be added, and inverted indexes created. Addition of fields within an existing group requires a reorganization. New data bases can be added any time without effects.
TOTAL	Automatic restructuring (adding/deleting a new data element, adding/deleting another level of hierarchy, and adding/deleting a repeating group field). Physical structural changes require redefinition, unloading, and reloading the affected files

2.5.3 Data updates. Updating is the process of adding, changing or deleting the contents of parts of the data base. Updating excludes restructuring of the data which requires a modification to the existing data definition or the logical data structures. Entering of new data differs in practice, from bulk data loading in terms of volume, mode, and input medium. For self-contained systems, there is usually a separate language for doing updates. The language may be somewhat analogous to interrogation in that some part of the data base must first be selected. Once the part is selected, it is updated rather than displayed. The update values may sometimes be specified as the results of a computation.

	Update Language Form	Update Values Via Computation
ADABAS	Analogous to query language	Yes
IMS/VS	Analogous to query language for GIS or HOST LANGUAGE CALL	Yes
INQUIRE	" " "	Yes
MODEL 204	Analogous to query language	Yes
S2000	Analogous to query language	Yes
TOTAL	Host language call	Yes

2.5.4 Update Levels. When a condition is satisfied and an update action is to be carried out on an entry, then the action may be on several levels. Updates can take place on the data element level. Items within a repeating group may be updated, or one update command may trigger actions applied to all records in the file. These system triggered updates include adjusting the pointers. If the action is a delete on a hierarchical structure, a valid consideration is whether the system automatically deletes all of the descendants.

	Data Element Level	Item within Repeating Group	Record Level	System Triggered Updates
ADABAS	Yes	Yes	Yes	All pointers adjusted by the system
IMS/VS	Yes	Yes	Yes	All pointers adjusted by the system
INQUIRE	Yes	Yes	Yes	No
MODEL 204	Yes	Yes	Yes	All pointers adjusted by the system
S2000	Yes	Yes	Yes	Depends on control command, e.g., REMOVE only removes data at a single level, REMOVE TREE removes the data and the lower level dependents. All pointers adjusted by the system
TOTAL	Yes	Yes	Yes	All pointers adjusted by the System

2.5.5 Other Maintenance Capabilities. Some of the global functions provided by the data base management system are checking authorization for updating and producing a daily update transaction for recovery, monitoring, or rolling back purposes. If on-line updates are allowed, a relevant consideration is whether the system locks out retrieval to the data base while it is being updated. Some of those functions can be imposed by setting up administration control.

	Check Authorization for Updates	Transaction Logged Facilities	Lockout for On-line Updating
ADABAS	Yes	Yes	Yes, at record level. Also, lockout only on updating functions and not retrievals
IMS/VS	Yes, via terminal identification and password	Yes	Yes, at segment occurrence level
INQUIRE	Yes	Yes	Yes, at record level
MODEL 204	Yes	Yes	Yes, at record level
S2000	Yes	Yes, via audit trial command	Yes, at the command level for multi-user update
TOTAL	Yes, via user supplied programs	Yes, via user supplied programs	Yes, at record level (up to four file levels)

2.6 Query Specification

2.6.1 Overall User-system Interaction and Search Specification. Query is the selection of data from a data base. Information is usually entered initially to identify the part of the data base on which the interrogation is to be performed. This is followed by a statement of the selection criteria and a statement of the action such as display or print. The query specification can be quite complex with the simplest case having one selection criterion and one action. The selection criteria are usually in the form of a Boolean combination (AND, OR, NOT) of simple conditional expressions. These expressions are in the form of triplets (data name, relational operator, and value). The form of the language covers the spectrum: English or English-like phrasing, system-forced dialogue and almost procedure-oriented programming with control transfers (labels and GOTO's).

	Language Form	Conditional Expressions	Logical Connections
ADABAS	English-like phrasing	Triplets	AND, OR, NOT, BUT NOT (with nesting)
IMS/VS with GIS	English-like phrasing	Triplets	AND, OR (with nesting)
INQUIRE	English-like phrasing	Triplets	AND, OR, NOT LINK (matching parallel repeating group) (with nesting)
MODEL 204	English-like phrasing with control transfers	Triplets	AND, OR, NOT, NOR
S2000	English-like phrasing	Triplets	AND, OR, NOT (with nesting)
TOTAL	Via Host Language		

2.6.2 Qualifications on Searches. Besides the logically connected conditions within the query specification, various qualifications may be imposed to further narrow the search. These are as follows:

Range searching -- User may specify upper and lower values and retrieve data whose value falls within the range.

Presence/Absence - User may retrieve records based on the existence or non-existence of a data field. In other words, null value is distinguished from zeros or blanks.

Condition on Groups - User may specify a conditions on the nth entry within a repeating group, a count of the number of entries within a group, or satisfaction of a special condition among the group entries.

String Matching -- For alphanumeric data, user may specify a literal string with some "don't care" characters or a partial string.

Phrase Searching - Search terms can be multiple words occurring within a text field.

	Range Searching	Present/ Absent	Repeating Group Cond.	String Matching	Phrase Searching
ADABAS	Yes, with reserved word THRU	No	Yes, thru positional subscripting	Yes	Yes
IMS/VS with GIS	Yes	Yes	Yes, thru positional subscripting	Yes	No
INQUIRE	Yes	Yes	Yes	Yes	Yes
MODEL 204	Yes	Yes	No, on nth entry but can be tested for value	Yes	Yes
S2000	Yes, with reserved word SPAN	Yes (EXIST/ FAIL)	Yes (HAS, AT)	Yes, with previously defined process	No
TOTAL	Via host language				

2.6.3 Aids on Search Formulation. During search formulation, most data base management systems require the retriever to know the precise name of data elements defined for an application, as well as which fields are keyed elements. However, some "forgiving" features may be built in to provide flexibility in the specifications of search criteria. These features include:

Phonetics -- The system will support a list of same sounding but differently spelled item values so that if the exact spelling of a search descriptor is not known, a phonetic match can be searched.

Synonyms - The system will support a synonym table or a controlled vocabulary so that the retriever may look for a class of semantically equivalent terms. This feature is useful in a text-oriented application.

Display of Data Element Dictionary - The system can display the data element dictionary if requested. The critical information during search formulation is the precise spelling of data element names and what data elements inverted for quick searching.

Natural English -- The system can "understand" a variety of input sentences in natural English. For example: the system allows reserved words or noise words such as IS, ELSE, BY, etc., for making the statements easier to read. Also, some systems permit grammatical alternatives such as, HAS, HAVING, thus making the statements easier to read.

	Phonetics	Synonyms	Display of Data Element Dictionary On-Line	Natural English
ADABAS	Yes	No	No	Yes
IMS/VS with GIS	No, but a separate alpha-search package can be used	No	No	Yes
INQUIRE	No	Yes, with optional MFL (multi-file) and DECON (data element control)facilities	Yes	Yes
MODEL 204	Yes	No, but may be implemented through invisible fields.	Yes	Yes
S2000	No, but "SPAN ALPHA" searches can achieve the same capability	No	Yes	Yes
TOTAL	N.A.	N.A.	N.A.	N.A.

2.6.4 Tutorial Features. Many systems provide tutorial features based on the assumptions that the retriever (user) may not be a programmer; e.g., information specialist, managers, or clerical personnel. This is particularly useful in the interactive query and search environment when the user might have at his disposal the following kinds of facilities:

"HELP" command - "HELP" or "EXPLAIN" commands which give assistance to a user on what options are available to him at various points in a session.

Documentation on-line - The system provides documentation on-line which display the relevant parts of documentation required by the user.

Browsing - This feature is especially useful in a bibliographic application or any retrieval producing a very large text output. Here the user may selectively browse through the search results by having an abstract or descriptive portion of the record displayed.

NOTE: All of these tutorial features can be achieved via predefined query or application programs.

	"HELP"	Documentation/ On-Line	Browsing
ADABAS	No	No	No
IMS/VS	N.A.	N.A.	N.A.
INQUIRE	No	No	Yes, with MFL (multi-file) option
MODEL 204	No	No	Yes
S2000	No	No	Yes, with "AT" command
TOTAL	N.A.	N.A.	N.A.

2.6.5 Multi-file Searching. A system may be designed so that for a single interrogation, data elements defined within a single file or a set of files may be retrieved. Multi-file searching may be achieved through a group of retrieval statements rather than a single statement. If the system operates on a set of files, the set may be coordinated by a common key-field or a linked pointer.

	Multi-file Searching	What Mechanism
ADABAS	Yes	Two files are coupled by a common linked pointer field via a utility program. At retrieval time, a single interrogation may specify conditions from coupled files but display the data elements in one file only.
IMS/VS	Yes	Via application programs
INQUIRE	Yes	Files are linked by any combination of inverted fields with linkage defined when the data bases are opened, or as part of the search statement. Up to 32 data bases can be handled simultaneously in a single query
MODEL 204	Yes	Via a group of retrieval statements. User specified linking by the data content in the files
S2000	Yes	The "HAS" operator can be used to logically connect files in one data base. Retrieval can be from one or both files
TOTAL	Yes	Via group of retrieval statements also linked pointers.

2.6.6 Predefined Queries for Future Invocation. Some systems have the capability to store frequently used queries. These queries may then be invoked by calling them by name. Implicit or explicit modification capability through the use of parameters may also be available in some systems.

	Predefined Queries
ADABAS	No
IMS/VS	Available in GIS
INQUIRE	Yes, there is a macro command system with conditional logic, prompting and parameters substitution
MODEL 204	Yes, with parameter list substitution capability
S2000	Yes, with parameter list substitution capability called "user-defined functions", strings, or in-line functions
TOTAL	Via host language

2.7 Output and Report Generation

2.7.1 Language Type. Every system provides some sort of output facilities. The facilities range from standard output provided by the system to sophisticated report writing capabilities. Most systems have both standard output for use during an interactive retrieval session, and bulk quantity report production facilities to be run as needed. The output specification language can be:

Standard Output - User specified WRITE or PRINT, followed by a list of data element names and system outputs in its own format.

Select Among Options - For some systems the output specification is guided by responding to a question such as "How many would you like to see" or "Which media do you want for your output".

Own Code - The user must write a program to specify the output format.

	Standard Output	Select Among Options	Own Code
ADABAS	Yes, with the DISPLAY command within ADASCRIPT	No	Yes, with host language system
IMS/VS	Available in GIS with LIST command	Available in GIS with LIMIT command can specify maximum number of output lines desired	Yes, with host language system
INQUIRE	Yes, name of field and value	User can preface the query with the limitation of the number of items retrieved	Yes, PL/1 or Assembler to INQUIRE tables, or with the host language
MODEL 204	Yes, with PRINT command	No	Yes, can be specified in user language and host language
S2000	Yes, with the PRINT or LIST command	No, but user can further refine searches through "DITTO" and "SAME" operators	Yes, with procedure language interface or with an optional Report Writer package
TOTAL	No	No	Yes, with host language

2.7.2 Media Flexibility. User may have control over the types of output devices. If the requestor is using a teletypewriter, the system usually responds on the same teletypewriter. However, for lengthy output, the user may specify off-line printing. Some systems provide output graphic display on special graphics terminals

	Off-Line Printing	Graphic Display	Other Special Output
ADABAS	Yes	No, but can be interfaced with an existing report writer	No
IMS/VS	Yes	No	No
INQUIRE	Yes	Yes	Linotron, tape, photo-composition and plotting packages
MODEL 204	Yes	No	No
S2000	Yes	Yes, tape output to Calcomp or other graphical display devices	No
TOTAL	System dependent not directly supported by TOTAL	No	No

2.7.3 Arithmetic and Statistical Capabilities. Computations on retrieved numeric data values are provided by the data management systems in a variety of forms. These computations are usually built-in system functions. The most common one are Totals, Averages, Count, Maximum, and Minimum. Some systems provide statistical functions such as Variance, Percentage, etc.

	Arithmetic and Statistical Capabilities
ADABAS	None available but can be interfaced to an existing report writer
IMS/VS	None available but can be interfaced to GIS, or use Host language capabilities
INQUIRE	All arithmetic operators and ABS, AVG, COUNT, DISTR (Distribution), MAX, MAXI, MIN, MINI, MOD, NUMBER, SELECT, SUM, DATE DIFFERENCES, STD, DEV.
MODEL 204	Count, add +, substract -, multiply *, divide / ,MOD.
S2000	All arithmetic operators and COUNT, MAX, MIN, SUM, AVE, SIGMA, TIME, DATE,
TOTAL	Host Language capabilities

2.7.4 Report Format Specification. Facilities for formatting reports cover a wide spectrum from complete user control to very little control. The amount of user control on the following functions is described:

Report Body Formatting - The specification of report blocks both vertically and horizontally across the paper.

Title and Footing Lines - The specification of title lines and/or footing lines to be included on each report page.

Column Control - The specification of column headings plus the ability to display the content of retrieved data in columnar structure with left/right justification.

Item Level Editing - The specification of "PICTURE" forms to allow for placement of decimal point, commas, insertion of leading or trailing characters, and leading zero suppression may also be specified.

Line Spacing - The specification of the number of blank lines on a page.

Automatic Page Numbering - Automatic numbering of pages from a specified initial setting.

	Report Body Formatting	Title and Footings	Column Control	Item Editing	Line Spacing	Auto Page Numbering
ADABAS	No (Yes with ADAWRITER)	No (Yes with ADAWRITER)	Yes (Yes with ADAWRITER)	No (Yes with ADAWRITER)	No (Yes with ADAWRITER)	No (Yes with ADAWRITER)
IMS/VS	No (Yes with GIS)	No (Yes with GIS)	No (Yes with GIS)	No (Yes with GIS)	No (Yes with GIS)	No (Yes with GIS)
INQUIRE	Yes	Yes	Yes	Yes	Yes	Yes
MODEL 204	Yes	Yes	Yes	No	Yes	Yes
S2000	No (Yes with Report Writer)	Yes (Yes with Report Writer)	Yes (Yes with Report Writer)	Yes (Yes with Report Writer)	Yes (Yes with Report Writer)	Yes (Yes with Report Writer)
TOTAL	No (Yes with SOCRATES)	No (Yes with SOCRATES)	No (Yes with SOCRATES)	No (Yes with SOCRATES)	No (Yes with SOCRATES)	No (Yes with SOCRATES)

2.7.5 Sorting. Sorting on the selected, retrieved data is usually provided by some systems usually in the batch mode. Sometimes the sort function is performed by a utility program external to the data management system, but operating under the operating system as a library routine. The specification of a sort usually includes the data element names to be used as sort keys, the sequence for sorting (ascending-descending) and the user's choice of collating sequence. The system may also have some requirements on the number of sort key fields and whether these fields are keyed or non-keyed.

	On-line Sorting	Number of Sort Keys	Ascending/ Descending	User Specified Collating Sequence
ADABAS	Yes	3 fields	Ascending only	No
IMS/VS	No	Available in GIS	Yes, both available in GIS	No
INQUIRE	Yes	60 fields	Yes, both	No
MODEL 204	Yes	Any number of keyed or non-keyed fields	Yes, both	Yes, the collating sequence may be specified as left-adjusted, string or numerical.
S2000	Yes	Any number of keyed or non-keyed fields	Yes (HIGH/LOW)	No
TOTAL	Host Language capabilities			

2.7.6 Extraction of Data. User may specify conditions such that retrieval data may form a subset of the original data base. The extracted data may be for the system's own use or may be output in machine-readable form for use outside of the system. There are certain restrictions on the format of the extracted data, such as it must be a COBOL file or a FORTRAN file or OS/360 file, etc.

	Extraction of Data and Its Format
ADABAS	Yes, user can write out the selected data to a tape via a procedure language (COBOL, FORTRAN, or PL/1) in a specified format.
IMS/VS	Available in GIS or Host Language, GIS allows creation of up to 99 "Temp" file in a specified format.
INQUIRE	Yes, there is an ability to create a temporary data base which may be searched or used as a link into other data bases. Can also extract data to any output medium in any user specified format.
MODEL 204	Yes, user can write out selected data creating a 360/370 file.
S2000	Yes, in 80 column card images in BCD via Immediate Access and Queue module or through procedural language in any format.
TOTAL	Host Language capabilities

2.8 Security Features

2.8.1 Protection Level. Most systems permit the user to define levels of security. These may be on the total data base, on the file level, or on the data element level. Physical security is also considered in terms of access to specific terminals which may be in a secure area.

	Data Base Level	File Level	Data Element Level	Terminal
ADABAS	Yes	Yes	Yes	No
IMS/VS	Yes	(Security provided to "segment" level) for IMS (GIS however goes down to data element level)		Yes
INQUIRE	Yes	Yes	Yes	Yes
MODEL 204	Yes	Yes	No	Yes
S2000	Yes	Yes	Yes	No
TOTAL	May be accomplished via Host Language routines			

2.8.2 Type of Authorization. Security control may be different with each user. For example, the system needs to distinguish between the user who has no access rights and the user who has rights to read and modify the data. Some systems can distinguish between users who are allowed to update data but not to restructure the data.

	Type of Authorization
ADABAS	15 levels of data security may be assigned. User password is translated to an access-level-number and an update-level-number. If the data element accessed has a higher level number than the user's access-level-number then this data element is locked to that user.
IMS/VS	A Program Specification Block (PSB) must be coded for every program that is written to access a data base. The code will indicate which segments the program will have access to and whether he may read, update, insert, delete or all of them.
INQUIRE	File level security provided by password system which restricts users based on the command being processed. Additional levels of security are incorporated in the security exit, which restricts access to the files, records and/or fields, based on the following parameters: password supplied by user; OS Job Name; Terminal ID; Command/Subcommand being accessed; Field(s) being accessed; Record(s) being accessed.
MODEL 204	The various types of authorization are: 1) updating data with requests, 2) updating pre-defined procedures (SEGMENTS) including editing, defining, and deleting, 3) updating data with SEGMENTS, 4) retrieving data with requests, 5) displaying SEGMENTS, 6) retrieving data with SEGMENTS.
S2000	20 passwords may be assigned, each having its own independent authorities for R(read), U(update), V(qualification for retrieval), W (qualification for update). Master password controls data base restructuring and certain data base management commands, including password/authorization maintenance.
TOTAL	May be accomplished via Host Language routines.

2.8.3 Backup/Recovery/Restart. In terms of protection against system failure due to physical equipment or software error, most systems provide some sort of backup, recovery procedure and restart mechanism. These might be, provision of periodical dump for recovery purposes, transaction logging for the recovery of data bases, etc.

	Backup, Recovery, and Restart Mechanism
ADABAS	System has log tape for backup and recovery. System also keeps transaction on a working data set during a session in case of system failure. This restart is automatic.
IMS/VS	Data recovery and transaction recovery through utilities provided with the package.
INQUIRE	System recovery achieved by utility program to copy log tape information on to the data base.
MODEL 204	System has a check point restart capability where physical record changes are kept in the log file. In addition, an audit trail provides a listing of all transactions which were made to the data bases.
S2000	The system has back-up files on tapes. Recovery can be invoked through basic language or through COBOL, FORTRAN, etc.
TOTAL	System has a log tape for back-up and recovery. Restart is up to user. Automated restart facilities are available with ENVIRON/1.

2.9 External Linkages

2.9.1 Procedural Language Interface. External programming interface is the ability of a system to provide attachment points for user written code. Procedure language interface is always provided in a host language type of data base management system. However, the facilities may also be available in self-contained systems. The interface usually consists of embedding high-level calls within a procedure language such as FORTRAN, COBOL, PL/1 or assembly language. These procedural languages supported are main tools for writing application programs.

	Procedure Language Interface
ADABAS	Yes (FORTRAN, COBOL, PL/1, Assembly)
IMS/VS	Yes (COBOL, PL/1, Assembly)
INQUIRE	Yes (FORTRAN, COBOL, PL/1, Assembly) These procedural language interfaces are all available in one package at a price.
MODEL 204	Yes (FORTRAN, COBOL, PL/1, Assembly)
S2000	Yes (FORTRAN, COBOL, PL/1, Assembly) These are all optionally priced packages.
TOTAL	Yes (FORTRAN, COBOL, PL/1, Assembly)

3.0 GOVERNMENT USER EXPERIENCE

Personal interviews were conducted with two Federal agency users of each of the six DBMS described in this report, except for ADABAS. Only one Federal government user of this system is known at present. The purpose of these interviews was to assess the effectiveness of each package from the user's viewpoint. Detail information such as volume of processing and sizes of data bases are cited only as a guide to support the overall performance of the DBMS. These figures quoted are current up to April 1975.

The interviews were oriented around the topics outlined below.

When Installed - The date when the DBMS was installed on the agency's own computer system. When a time-sharing service was used, the initial service date is given.

Background - This section provides the information about the agency's main mission and activities in the data processing context. Considerations as to why a particular DBMS was chosen are also covered here.

Applications - In cases where several applications are installed on one DBMS system the general characteristics of the applications are described together with some specific information concerning each application.

Computer Environment - The computer main frame and the operating system environment in which the DBMS resides are covered. If the system is used on-line, the number and (sometimes) the types of terminals used are presented here.

Operation and Use - The information includes the volume of data in terms of on-line queries, on-line updates, batch queries, and batch updates.

Application Data Characteristics - The sizes of the data bases and the types and structures of data defined for each application are covered in this section. In some cases, the names of data elements representative of the application are given as examples.

Application Program Characteristics - The types of application programs and the languages used to implement them are discussed. The manpower requirements and the skill levels necessary to support the application programs are also described.

Users - The types of users who perform retrieval and update operations are covered.

Experiences with DBMS - This section covers the effort required to install the DBMS system. Data definition and loading, vendor support in terms of training, fixing problems, and system enhancement through user group activities are also covered in this section. A subjective opinion is expressed with regard to the documentation of the DBMS.

Overall Impression - This section covers opinions and impressions by government users, as interpreted by NBS. System reliability, recovery procedures, and how the system responds to changing data structure requirements are covered. The acceptability of each system is described in terms of user impressions.

3.1 ADABAS Used at U.S. Navy Fleet Material Support Office

Name of Agency: U.S. Navy Fleet Material Support Office (USNFMSO)

DBMS Package Used: ADABAS

When Installed: In November 1974 an initial contract was drawn with Planning Research Corporation (PRC) to use ADABAS on the PRC time-sharing service. Since January 1975, USNFMSO uses its own terminals to query its ADABAS application data base via telephone lines.

Background: The U.S. Navy Fleet Material Support Office provides information about aviation maintenance activities. It collects and disseminates maintenance data on aircraft and ship materials.

An evaluation of generalized data base management systems was done prior to the acquisition of ADABAS. Other packages considered where IMS, TOTAL, and SYSTEM 2000. ADABAS was chosen because of its "blackbox" approach, i.e., the user never needs to know "how" or "where" his data is stored. Adding and deleting data fields can be achieved without unloading and reloading the data base. USNFMSO will have a large data base (over 3.3 billion bytes) and data compression is a mandatory requirement. Also ADABAS is available on the PRC time-sharing service, so that a prototype operation could be carried out without acquiring and installing the package at USNFMSO.

Application: At present, USNFMSO has one application on ADABAS: the Navy Maintenance Data Base. This data base consists of information on aircraft components and maintenance activities (date, man-hours worked, parts replaced, etc.)

Computer Environment: ADABAS is installed on PRC's IBM 370/155 with TSO. PRC provides USNFMSO with its own copy of executable ADABAS which uses about 210K bytes of core. USNFMSO also owns two mountable disk packs which are maintained and mounted at PRC. At present there are six terminals located at Mechanicsburg, Pennsylvania; it is estimated that this number will increase to about 60 terminals.

Operation and Use: On-line querying is the primary use of ADABAS at USNFMSO. Typical queries are requests for counts of items satisfying some Boolean combinations. Numerical comparision on ranges is common within a query. The number of queries per day is still unknown at this time.

No on-line updating is done. Data to be entered or modified are received monthly and an update run in batch is made approximately once a month.

Application Data Characteristics: The Navy Maintenance Data Base at present has 50 million bytes residing on two 3330 disks at PRC. The Navy will soon be installing two more 3330 disks. The data base consists of 13 separate files with an average record size of about 100 bytes. The average number of data elements in a file ranges from 12 to 75 fields. Hierarchical and repeating groups are used. The data is 65% numeric which is stored in binary form. The remainder consists of alphanumeric and coded data which are compressed by ADABAS. ADABAS reportedly has compressed the data base by 30%.

Application Program Characteristics: Although ADABAS can be interfaced to a batch report writer (ADAWRITER, also marketed by Software Ag), USNFMSO does not have it but is working on an interactive report writer callable from the ADASCRIPT language. This interactive report writer is being written in assembly language and will produce customized reports and perform extensive arithmetic computations. A batch update program is also being planned. It will be written in COBOL to suit the existing tape format. No other application programs are necessary at this time.

Users: The users for on-line querying are mostly upper level managers or information analysts. ADASCRIPT, the query language for ADABAS, is used for conditional retrieval and display. USNFMSO personnel include one analyst and three software specialists for supporting application programs development and batch updating of the data base.

Experiences with ADABAS: USNFMSO did not install ADABAS on its own machine because its IBM 360/65 is not configured for real time teleprocessing applications. A contract was negotiated with PRC to use ADABAS through the time-sharing service as a feasibility study.

The software specialists, after attending ADABAS' one-week training course and with some assistance from PRC, were able to define the data for loading which was done in three hours. Later refinements to data definitions (changing from keyed to non-keyed or vice-versa) can be done by running an ADABAS utility program. No unloading and reloading of the data base is necessary.

Vendor support is reported very good. Constant enhancement is being made to ADABAS. For example, a revised version of ADABAS due July 1975 will increase the throughput by 50%. The documentation is acceptable. USNFMSO have found that more information is needed in the documentation in certain areas but the vendor is very responsive in helping the users.

Overall Impression: USNFMSO considers ADABAS a very cost-effective package. "You get what you paid for." USNFMSO maintains a large scale data base and data compression is very important to them. The data elements needed for search queries are usually unforeseen; therefore, it is difficult to decide at data definition time which data elements should be key fields. In other generalized systems, accessing a field which is not keyed involves a sequential search, or prohibits a search; ADABAS maintains an address map (an "Association") so that search of a non-keyed data element is limited to a subset of the data base using the "Association" address. ADABAS data bases never need to be reorganized, and very little application programming is required to operate ADABAS. USNFMSO feels that they can fully utilize ADABAS without extensive training or detailed knowledge of the system.

ADASCRIPT is free-form and easy to use, but does not provide enough of the arithmetic computation capabilities and report formatting required. (These features are all available with ADAWRITER.) However, ADASCRIPT is non-proprietary and USNFMSO finds it very easy to build its own application programs and interface to ADASCRIPT.

The package is reported to be extremely reliable and has caused no problems.

3.2 IMS Used at Naval Materiel Command Support Activity

Name of Agency: Naval Materiel Command Support Activity (NMCSA)

DMBS Package Used: IMS

When Installed: Fall 1970.

Background: NMCSA performs logistic type activities for such functions as personnel, finances, weapons systems analysis, aircraft maintenance, military construction, and ship alteration management. Batch and teleprocessing capabilities are required, and the Navy needed a system that could be installed quickly. The first application of IMS-1 with IMS/data communication was a small financial system; it was operational within a few months. NMCSA has since upgraded to IMS/II.

Applications: There are about 40 IMS applications at NMCSA. The largest of these is the Integrated Financial System which contains information on obligations, commitments, negotiations and contract monitoring details. The agency utilizes the IMS multi-file capability to logically connect several data bases. These include a general ledger data base, a document history data base, a suspense data base, an accounts data base, and a document status data base.

Computer Environment: IMS is operating at NMCSA on an IBM 360/65 under OS/MVT with IBM 3330 and Itel 7330 disk packs. The IMS nucleus requires 530,000 bytes of core. All of the available access methods are utilized. There are 74 communications lines over 120 terminals located throughout the country, with the majority of the terminals located in the Washington, D.C. area.

Operation and Use: NMCSA does most of its file updates in batch mode, although there is an on-line update capability. Update is restricted to a few terminals. There is slightly more query activity than update activity. Ad hoc queries are via QL1 (a query language developed by COMRESS) which runs as an IMS application program. Predefined queries are written in IMS-COBOL. A simple query takes a few seconds; a complex query can take up to two minutes. (These performance reports may reflect the saturation rather than IMS operational efficiency.)

Application Data Characteristics: There are more than 40 separate physical data bases, logically connected via IMS. They range in size from 15 million characters to 180 million characters. The Integrated Financial System

on-line is estimated to reach 500 million characters. Textual data, coded information, and numeric data are all included in this integrated system with 3 to 4 levels of hierarchy.

Application Program Characteristics: IMS requires application programs for such functions as query, data validation, and report writing. At NMCSA, COBOL is the primary application programming language. Although the IQF (Interactive Query Facility) query language is available as a separate IMS module from IBM, NMCSA elected instead to purchase QL1 from COMRESS. QL1 is written in assembly language and PL/1.

Extensive requirements for arithmetic computation on financial data are met by in-house-developed COBOL programs. Application modules which can be invoked by transaction code with user-supplied parameters facilitate preformatted retrieval and display of reports such as those required by the Navy Comptroller. Since IMS provides security only to the segment level, application programs had to be written to perform security checks at the data element level.

Users: The end users of the Integrated Financial System are accounting clerks, military personnel, project managers, and secretaries. In-house training is provided to these users. A user guide is also available. It was written by NMCSA programmers/analysts and consists of pre-formatted transaction calls.

Experience with IMS: There is a good IMS community for trading information--the IMS SHARE/GUIDE Group. The Washington area users group meets once a month. Recent vendor support and documentation are reported to be better than they were when IMS was first marketed.

Overall Impression: The IMS capability to support multiple applications in an integrated data base environment, in both batch and on-line operations, satisfied NMCSA's requirements. The need for extensive in-house application programming and the complexity of IMS are compensated for by the flexibility in data organization and administration offered by the system.

3.3 IMS Used at the Federal Reserve Board

Name of Agency: Federal Reserve Board (FRB)

DBMS Package Used: IMS

When Installed: March 1973

Background: In 1972 the FRB did extensive benchmark testing of two systems, TOTAL and IMS, on their data base of bank information. The systems were evaluated against a group of desired capabilities, and both systems successfully completed all tests. When relative weighting factors were applied to the various capabilities, with heavy weights assigned to flexibility in data base structure and definition as well as data administration support, the decision was made to lease IMS from IBM. IMS/DB (Data Base) was acquired; IMS/DC (Data Communications) will be installed at FRB in July 1975 to permit terminal access.

Applications: The Bank and Statistics Data Base is the largest IMS application at FRB. It is an integrated system of several separate physical data bases: current financial data on bank branches; historical financial data on bank branches; and current and historical bank structure information. These data bases are logically connected via application programs bi-directionally for analysis and maintenance.

The FRB is building another application data base at a higher level in its hierarchical file structure, for information on bank holding companies. This will be logically connected to the Bank and Statistics Data Base.

Another application file is the IMS Data Dictionary which stores data administration information such as validity and quality edits. This file is terminal oriented and occupies 3,000 tracks of disk.

Computer Environment: IMS/DB is installed on the FRB IBM 370/168 with 2 megabytes of core under OS with TSO. VS-2 is being installed in March 1975. IMS/DB can be run under TSO as a batch job.

Operation and Use: File update is mostly in the batch mode from cards. Bank structure data is updated daily as needed. Much of the financial data is updated querterly when the bank reports come into FRB. FRB is planning to do on-line updating when IMS/DC is installed.

There is a heavy volume of retrieval operations. There are 35 to 40 terminal users at a time at FRB, with 5 to 10 of these on IMS. Twenty-five to fifty queries per day are run on the IMS files. The volume of data entry is small, but much data revision is performed.

Application Data Characteristics: The Bank and Statistics Data Base contains one record for each of the 17,000 banks and 46,000 bank branches in the Federal Reserve system. The bank records go back to 1960, and the branch records go back to over four years. The data now occupy more than one 3330 disk pack. It is projected that two such disk packs will be required. The data are organized hierarchically. The root segment on bank structure contains approximately 35 fields. Data elements of banks are characterized by type, size, location and name. Logical relationships permit access by bank market area. Historical research on banks, though complicated by bank mergers, is possible with the integrated data base. The records are a combination of textual and numeric information. The textual fields are fixed-length at present, but variable length fields are planned when VS-2 becomes operational.

Application Program Characteristics: FRB has stringent update and edit requirements and has written its own application programs to insure validity and quality checking on the data bases. These edit functions are stored outside the programs in the IMS Data Dictionary.

FRB programmers have written extensive application programs to interface IMS with the users. They have written query subroutines in ALC for one application. Their predominant application programming language is COBOL, although they have also written in PL/1, ALC and FORTRAN. There are two IMS support programmers and about 20 IMS application programmers. Recently, the data administration fuction has been separated from the application programming function.

Users: Querying of the IMS files is done mostly by FRB via application programs. File maintenance is performed by both data processing unit personnel and by some end-users.

Experience with IMS: IBM provides good on-site support for IMS. Three days of formal training were provided initially, but now IMS training is conducted in-house by FRB personnel. Documentation is reported to be extremely comprehensive.

Over 150 installations are members of the IMS Users Group (SHARE/GUIDE). FRB considers this group to be very valuable in that it facilitates communication between users and it provides a formal channel for requesting system enhancements.

Data base administrators and systems programmers from the various user installations meet regularly, and IBM's IMS specialists (not salesmen) are invited to participate. Recently eleven significant new enhancements recommended by the Users Group were announced as part of the latest release of the product.

Overall Impression: The FRB requirements for flexibility in data base structure and definition are satisfied by IMS. Although IMS is a very complex product requiring much application software support, it offers significant capabilities for data reorganization and redefinition. FRB considers data flexibility to be as important as the system's efficiency and resource utilization. Integrated data administration which permits queries on data that are logically related but physically independent is a capability of IMS that is heavily utilized by FRB. Changes to data relationships can be achieved via rule redefinition without requiring major program modification or data re-entry.

3.4 INQUIRE Used at the National Library of Medicine (NLM)

Name of Agency: Office of Computer and Communication Service
National Library of Medicine

DBMS Package Used: INQUIRE

When Installed: November 1972. The package was leased as per the GSA schedule, and will be fully owned by NLM in November 1975.

Background: NLM staff had been using TYMSHARE* services for software development and for data base applications other than ELHILL, the major NLM bibliographic medical literature system. It was considered too expensive to keep these application data bases on-line via TYMSHARE, so a DBMS was sought which could be installed on the NLM computer. On-line file maintenance and on-line Boolean search capabilities were required. INQUIRE was chosen.

Applications: Currently there are about 15 NLM applications on INQUIRE. The largest data base, INPROC (in-process) consists of accounting information for monographic material which is in the process of being acquired or catalogued. It is the NLM management tool for its library acquisition function. Similarly, AVPROC is the in-process file for audio-visual material.

*TYMSHARE is a commercial timesharing service.

SERLINE (serial records on-line) is a data base of 6,600 journals and contains information on which of the 117 libraries in the NLM network holds which journals. SERLINE is generated and maintained on INQUIRE which creates an input file for ELHILL.

EMPMIS (extramural program management information system) contains grants management data and also several files on grant information for resource, research, training and publication. Records are created from grant applications, and are retained regardless of subsequent acceptance or rejection of the applications.

MEDMIS (Medline network management information system) has several files which contain accounting, statistical, and administrative data for each Medlars/Medline user institution. For each institution, such information as training received, billing agent, addresses, log-on codes, etc. is included.

DOCLINE (document retrieval on-line) contains bibliographic information for inter-library loan activities.

Computer Environment: INQUIRE is installed on NLM's 370/158 under VS 2, release 2, with TSO. The NLM applications run with 168k bytes of core. Any teletype-compatible terminal recognized by TSO can be used. There are approximately 30 terminals, with a total of about 40 users, at NLM. There also are four communications lines from UCLA Medical School, Harvard Medical School, National Medical Audiovisual Center in Atlanta, and the British Lending Library through the ARPANET.

INQUIRE is not re-entrant. TSO is multiple threaded and processes many users at one time. At any single INQUIRE application there may be several accesses from different users.

Operation and Use: For the INPROC application INQUIRE is used approximately 600 hours per month with an average of seven terminals operational concurrently. In some applications, the volume of file maintenance is high, with on-line input, but deferred (overnight) file update. Periodic reports are generated in batch mode and/or on-line with a great use of pre-stored commands and/or report formats. Most of the on-line searches are utilized for management information requirements. While off-line searches are primarily utilized for report generating.

Application Data Characteristics: All the INQUIRE data bases at NLM are predominantly textual, with little if any coding of data. More than half of the INQUIRE use is on the INPROC data base which consists of 27,000 records, and is growing. Each INPROC record contains title, publisher/dealer, place of publication, requisition number if ordered, date received, etc. The estimated average record size is 920 bytes, the largest is 1,524 bytes.

Application Program Characteristics: NLM has built very few application programs for INQUIRE. PL/1 is the NLM application program language. Examples of application programs are a front-end conversion program for upper-lower case and Greek letters, and a post-processor to convert SERLINE output to ELHILL input format.

Users: There are only one full-time and four part-time INQUIRE data base administrator/system analysts at NLM. Most of the users directly use the INQUIRE language and these are professional librarians, management personnel, and library technicians.

Experience with INQUIRE: INQUIRE is easy to load on IBM equipment through the use of an IBM utility program called IEHMOVE. The IBM linkage editor is used to combine modules. At present, NLM keeps at least 5 copies of INQUIRE: basic INQUIRE on-line, basic INQUIRE in the background, and a basic INQUIRE in the background not overlayed. Associated with each of the first two copies is a statistical package and a high performance sort.

Data base definition is facilitated by a preprinted form. Data entry may be done on-line or in batch. On-line input is performed with interactive prompting for all or selected fields of the record to be added. Validation of the record is performed as prescribed in the data base definition. In background mode, the user may add records by executing INQUIRE itself or a series of programs called the Loader System (INQUIRE-provided software with the basic package). INQUIRE additions allow punched cards or TSO data sets that contain field names and field values. The Loader System allows data in this same (standard input) format or a "direct load" format. (A limitation on direct loading is that only the last field in each record may be of variable length.) NLM reports that it takes about two minutes of connect time to enter a new record with twelve fields with interactive prompting. The INQUIRE response time may be due to the NLM-imposed priority schedule which favors ELHILL rather than INQUIRE. Recently, INQUIRE use was restricted to non-peak hours. Use of a software monitor (Boole & Babbage) and the INQUIRE-provided accounting routines pointed up areas where changes in

file definition could improve performance.

Vendor suport, both in correcting problems and in providing enhancements, is good. There is a users group whereby users can propose enhancements to INFODATA. Documentation was reported to be improving. INQUIRE initially provided training and assistance in definition for one application, but now training is available at an additional cost. System reliability is good.

Overall Impression: NLM's requirement for an in-house DBMS with on-line maintenance and query is satisfied by INQUIRE which is relatively inexpensive, easy to install, and requires little special programming. INQUIRE is built for quick retrieval. The search technique utilizes three distinct files: Index File (last address of last key data elements), Search File (chained addresses of all occurrences of keyed values) and Data File (actual data values). The chained method facilitates quick searching but updates are not as straight forward. Some NLM applications are heavily oriented towards maintenance, there is general satisfaction because they do deferred updates.

3.5 INQUIRE Used at Law Enforcement Assistant Administration (LEAA)

Name of Agency: Law Enforcement Assistant Administration (LEAA)
Department of Justice

DBMS Used: INQUIRE

When Installed: 1972 under a lease agreement.

Background: Under contract, Boeing Computer Services, Inc. did a study for LEAA to find a DBMS which could satisfy LEAA needs for flexible report formatting and file maintenance. INQUIRE was chosen and leased by LEAA in 1972 for installation by Boeing.

Applications: There are five INQUIRE application files at LEAA, all concerned with grants management:

1. The Block File contains information on grants of blocks of money to states by LEAA. There is one record per state in this file and approximately six years' information, for a total of about 360 records.

2. The Direct Grants File, contains information on direct grants to individuals, institutions, etc. The file consists of 5,000 records with about 130 data fields per record.

3. The Subgrants File, the largest LEAA INQUIRE application, contains information on grants by the states from LEAA block money. This file has more than 60,000 records with about 2,500 characters per record.

4. The Application File provides the means for tracking grant applications through the LEAA grant process from receipt through acceptance or rejection.

5. The Query File is the record of queries made on the above four files. It is maintained for management purposes.

Computer Environment: Since the Department of Justice computers do not have TSO, LEAA has contracted with Boeing Computer Services, Inc., in McLean, Virginia to have their own copy of INQUIRE operating on the Boeing IBM 370/168 under VS-2 with TSO. Private LEAA disks are maintained at Boeing, two of which are on-line at all times. LEAA accesses INQUIRE via terminals (Hazeltines, 2741's, Anderson-Jacobson, etc.) There is delivery service twice a day for line-printer output.

Operation and Use: Ninety-nine percent of the data coding, input and maintenance is done by Boeing for LEAA in the batch mode. On-line update is done at LEAA terminals only for minor data correction. Ninety-five percent of the file queries are also done in the batch mode, but queries are always initiated on-line at LEAA. There are approximately 30 to 40 queries per day and one grouped update run per day.

Application Data Characteristics: The LEAA grants files contain much coded informaton for which the INQUIRE decode capability is utilized. Records can contain repeating groups of coded information on grant categories and objectives; these codes are primary search keys. Numeric information on grants is maintained, as well as project title, dates, status and location information. There is also a textual field, the project summary, in the direct grants file, which can contain up to 1,440 characters. This field is searchable sequentially, but is rarely searched.

Nothing in these files is classified since all the information is in the public domain. In order to protect the files, passwords are required for access.

Application Programming Characteristics: All application programming is done by Boeing. There are several pre-editing programs, written in COBOL, which validate dates, location codes, etc.

The INQUIRE macro facility is extensively used for report writing. These macros are stored in a Macro Library, and are invoked with parameters to produce customized reports.

Users: LEAA responds to queries from Justice, Congress, and the public at large. Sometimes a charge is levied for services. There are four LEAA staff members who query the files, one data base administrator, and three subject specialists.

Experience with INQUIRE: In accordance with its maintenance contract with LEAA, INFODATA gives assistance in data definition and loading. For example, it has helped LEAA analyze the tradeoffs between keyed and sequential access during data definition. A one-week training course was provided as part of the original installation contract. Now LEAA staff are trained in-house. Documentation is described as complete, but sometimes difficult to understand. Recently a combined index to the user manuals was provided by INFODATA in order to improve access to the manuals.

LEAA now uses the new INQUIRE multi-file capability, and finds it satisfactory. A relationship between files must have been established before use of the multi-file capability.

Overall Impression: LEAA's requirement for flexible report generation is satisfied by INQUIRE's macro, decode and sorting capabilities. Very little special programming has been needed. Computer system problems are expeditiously handled by Boeing. INFODATA support on INQUIRE problems is described as very responsive. File update is considered quite expensive in terms of computer time. The basic INQUIRE system is easy to learn, but sophisticated features require extensive experience.

3.6 Model 204 Used at The Department of State

Name of Agency: Department of State

DBMS Package Used: Model 204

When Installed: Leased with purchase option for the State Department in September 1974 at a cost of $2,070 per month for 36 months.

Background: The State Department has been involved with data base applications since 1966. They have developed several in-house software systems: MYSTIQOL, a sequential access generalized data base management system; DISPLAY, a report writer; FOCUS, a data base editor and updater; and IQ, a batch query language. The State Department needed a data base management system which has all the features in their present in-house systems plus additional features, such as inverted fields, storage compaction, processing speed, and multi-file capability. Some of the mandatory requirements specified in an RFP sent out in April 1974 were:

- Capability of supporting a data base of six billion bytes
- A checkpoint capability
- Accessibility to programs written in FORTRAN, COBOL, PL/1 and ALC
- Security protection on the file level and also on the record/field level
- A sort capability
- A multi-server capability.

Model 204 was chosen as the lowest cost data base system satisfying all of these requirements.

Application: There are four applications which use Model 204.

1. A file of foreigners given grants to vist the United States. This is a small file with about 23,000 records, consisting of ten data fields per record.
2. A mailing label file of all persons who might receive mail from the State Department together with the categories of their interests. This file has 20,000 to 30,000 records with 20 to 30 data fields per record.
3. An experimental personnel file which will essentially contain personnel and payroll information about every employee of the State Department. This file will consist of 16,000 records with several hundred data fields.
4. The CERP (Current Economic Report Procedures) file. This is a small data base.

Computer Environment: Model 204 is operating on an IBM 370/158 under VS2 release 1.7 operating system. It uses less than 200K bytes of core. At present, the State Department is still reviewing a separate teleprocessing interface because the teleprocessing interface provided by Model 204 does not support IBM 1050 terminals at the State Department.

Operation and Use: Overall usage consists of more updates than queries at the present time. The updates, numbering about 200 per day are done in batch mode. On-line updates are planned for the mailing label file and the personnel file. At the time, the only on-line activity uses the FOCUS package as an input editor.

Application Data Characteristics: In general, data are fixed length. Coded data are heavily used. No hierarchical structure is required, but repeating group fields are used.

Application Program Characteristics: Input editing is accomplished via a COBOL program - FOCUS. Few application programs are necessary because the Model 204 user language is flexible enough to do all the required functions.

Users: The users are subject matter specialists who received three days of training in the user language of Model 204. They perform most of the querying of the data base and report writing. Updates to the data bases are done by programmer/analysts.

Experience with Model 204: Installation of the system takes 1 to 2 days. Data definition and loading require expertise with Model 204 because disk allocation space must be specified. The support from the vendor has been good, including adequate training for the users. The State Department users say that Model 204 documentation is quite complete, but could be better organized.

Overall Impression: The State Department is pleased with Model 204, claiming that the performance is extremely fast. (A test was performed retrieving 7,800 records from four files of 15,000 records each. This was accomplished at approximately six records per second.) The State Department would like to see two features added to Model 204. These features are: (1) direct handling of off-line storage such as tapes, and (2) a flexible format for data input editing. All the other features requested by the State Department are provided by Model 204.

3.7 Model 204 Used at Food and Drug Administration

Name of Agency: Food and Drug Administration (FDA)

DBMS Packages Used : Model 204

When Installed: March 1973.

Background: A request for proposal was issued in January 1973 for a generalized data base management system. Among the criteria used in system selection were these:

- Off-the-shelf, operational system for IBM 360/370
- Support for 60 concurrently operating low speed remote terminals
- Fail-soft error recovery procedures

- On-line and batch mode capabilities
- Concurrent accessibility of the same file by multiple users
- Ascending and descending sort capability
- Three levels of system security
- Simple data base creation
- End-user oriented with English-like retrieval language
- On-line report writing
- Data base directly accessible from ANS COBOL programs.

Model 204 was selected as meeting these criteria.

Applications: The FDA installation is a service center providing computing service to the Public Health Service. At the present time, there are over thirty different applications being supported on Model 204. There are application systems from the Alcohol, Drug Abuse and Mental Health Administration, Food and Drug Administration, (Bureau of Foods, Bureau of Radiological Health, Bureau of Medical Devices), National Institute for Occupational Safety and Health, Health Services Administration, and the Assistant Secretary of Health. The systems are mostly administrative in nature, consisting of small to medium size data bases (averaging 5,000 records of 200 to 300 characters per record). Examples of some of the files are: the Program Management Information System which contains information pertaining to all Federally funded drug programs; the Grants Data System which contains information about all active grant awards; a Drug Prevention Reporting Information System which contains an inventory of National Drug Treatment Programs. There are four personnel files, two sample personnel files used for instructional purposes, one personnel file for demonstration purposes and one file which maintains an on-line roster of Bureau of Food employees.

Computer Environment: Model 204 (FDA in-house designation of Model 204 is ADMIR - Administrative Management Information Retrieval) operates on an IBM 370/158 under OS/MVT. Approximately 450K bytes of core are used, supporting application files. IBM 2741 type terminals and teletype terminals (300 baud) are used with Model 204's teleprocessing interface program in a multi-thread environment whereby application programs may operate concurrently.

Operation and Use: The users of Model 204 operate predominantly in the on-line mode. There are applications that do on-line data entry or off-line data entry exclusively, and others that use a combination of both. Over all applications it is estimated that twice as many updates as on-line queries are performed monthly.

Application Data Characteristics: The files contain approximately 75% alphanumeric and 25% numeric data. Data are not hierarchical, but there are multiple field values. Application files vary in number of fields from very small (10) to very large (255) and in volume which ranges from 200 to over 10,000 records.

Application Program Characteristics: On large updates or initial preparation of the data base, COBOL, PL/1 and other procedural languages are used to edit the data and place it into an acceptable format. Procedural language can be used with IFAM (Model 204's Inverted File Access Method) to update files in a batch mode.

Users: The users of Model 204 applications are generally non-ADP specialists. The data bases are defined and created for the users by the FDA computer service center. The users maintain their data base and write queries and reports in the Model 204 user language.

Experiences with Model 204: Computer Corporation of America installed Model 204 for FDA. Any problems which arose were quickly corrected by the vendor. The documentation provided with the system included the source code in BAL. The documentation at first was not comprehensive but is improving. In two years of operation there have been very few Model 204 failures.

Overall Impression: FDA wanted a system that was already in use, one that required little support effort after initial file definition and creation, and one in which non-ADP users could operate with little training. Other criteria concerning the technical capabilities of the system were also met by Model 204. The vendor support has been good; performance of the system is also good. FDA has been very satisfied with Model 204.

3.8 SYSTEM 2000 Used at Department of Agriculture

Name of Agency: Department of Agriculture
Statistical Reporting Service (SRS)

DBMS Packages Used: System 2000

When Installed: In June 1974 SRS began use of SYSTEM 2000 via the INFONET Time Sharing Service, provided by Computer Sciences Corporation under GSA contract.

Background: At the time of system selection, System 2000 had department-wide approval as the Agriculture data base management system. System 2000 was accessible by SRS on the Department's computer at the Washington Computer Center, and also on the INFONET system. The INFONET timesharing service was chosen to develop the early stages of an eventual USDA network service, because it provided a national processing network which SRS required and which the Washington Computer Center did not provide. The application of System 2000 is still in the development stage.

Applications: The Official Estimates Data Base (OEDB) is a collection of data bases of the official published estimates for crops, livestock, farm labor, prices paid, and county estimates. Commodity statistics and official estimates are released by the Statistical Reporting Service. Last year, there were 589 releases; this amounts to approximately 50 reports per month.

Computer Environment: The INFONET Timesharing Service of Computer Sciences Corporation makes use of UNIVAC 1108 computer systems. The Computer Sciences' Timesharing System (CSTS) is a proprietary operating system developed by the Computer Sciences Corporation. INFONET provides a nationwide network service supporting concurrent interactive timesharing and remote batch processing.

Operation and Use: On-line queries are not presently being used because the OEDB is not operational. However, on-line queries are expected to be about 100 per day via the computer network. On-line updates are not planned for the completed system. All modifications to the data base will be batched under SRS control. Batch queries are used to generate the output reports. The volume of batch queries is estimated at about 15-30 per day.

Application Data Characteristics: Before release to the public, the information in the data base needs to be kept confidential. Two separate files -- public and private -- are being developed to provide the required security. The present developmental data base of crop information contains approximately 100,000 characters. The full OEDB will consist of six files: Private Crops, Public Crops, Private Livestock, Public Livestock, Private Prices Paid and Farm Labor, and Public Prices Paid and Farm Labor. The expected size of the Crops public file will be 2 million characters per year with a history of 5 to 10 years. There are 71 data elements for the Crops Public file containing six levels of hierarchical definition with

repeating groups. The data types are almost all numeric of fixed length.

Application Program Characteristics: Approximately six COBOL PLI (System 2000's Procedural Language Interface module) application programs will be developed for the crop data base. The functions are all for pre-processing the data prior to entering into the files. Further application programs such as utility programs to select, print and plot data are being planned.

Users: The envisioned use of the system is for direct support of the estimating process, and therefore the users will be subject matter specialists from various parts of the country who will query the data base. The maintenance of the data base will be carried out by programmer/analysts of SRS. SRS has at present two data base administrators plus a pool of about 26 programmer/analysts who will be supporting the OEDB application.

Experiences with System 2000: Since the INFONET system is a service, there was no need for system installation. The data definition effort required approximately three months of planning, and loading the definition into the system took about two days. Data loading for the System 2000 was accomplished by writing a COBOL program. INFONET provides consultation, documentation, and training to their subscribers.

Overall Impression - SRS feels that System 2000 adequately fits their requirements. The system not only has Agriculture Departmental-wide approval but also provides an easy to use language for querying. However, the unique security requirements, i.e., security at the data element value level, cannot be conveniently handled by System 2000.

3.9 SYSTEM 2000 Used at Environmental Protection Agency

Name of Agency: Environmental Protection Agency (EPA)
Management Information and Data Systems Division

DBMS Package Used: System 2000

When Installed: Purchased in March 1974.

Background: EPA wanted a data base management system that would operate on two different computer systems: the UNIVAC 1110 in North Carolina and the IBM 370 in Washington, D.C. In addition, they wanted a system that would be operational in a short time frame with various types of needed applications. The system would also need to be flexible in file maintenance and have an end-user oriented query language.

System 2000 was the system chosen to meet these criteria.

Applications: EPA has various applications on System 2000. One of these is the Air Pollution and Emissions Testing System. The Mobile Services Enforcement Division of State and Local Governments maintains information on the most polluting motor vehicles in use. The data base contains information on four different test methods and the results of each test method on different vehicles. Another application is the Spill Prevention Control and Countermeasures (SPCC) data base system. This system contains information about various facilities plans on spill prevention. It also stores descriptions of the facilities, circumstances of the spills, measures taken or contemplated, etc.

Computer Environment: System 2000 is operating on an IBM 370/158 at Optimum Systems, Inc., the contractor providing computer services to the EPA. The IBM version of System 2000 runs under TSO using 170K bytes of core in a single thread mode. System 2000 is also operational on the UNIVAC 1110. The applications described below are those on the IBM system.

Operation and Use: For the Air Pollution and Emissions Testing System, the majority of updates occurs in batch mode. Reports are also produced in the batch mode. On-line queries will be used at field sites. The volume of queries is not known at this time. At the present time, most of the effort is in constructing the data base. However, future expectations are for more on-line query activity from various field sites. For the SPCC application, there will be all batch updates and on-line entry of queries. Special or standard reports will be produced in deferred batched mode.

Application Data Characteristics: In the Air Pollution and Emissions Testing System, data are mostly numeric and have three levels of hierarchy with repeating groups. Approximately 200 million characters of data are stored in this single data base. In the SPCC application, data are mostly alphanumeric and have three levels of hierarchy with repeating groups. Each record contains eight data elements.

Application Program Characteristics: In the Air Pollution and Emissions Testing System, a COBOL program is used to edit the input data. A COBOL PLI (System 2000's Procedure Language Interface Module) program loads the data into the data base. A "Summary Retrieval Program" written in COBOL PLI extracts data from the data base and passes this data to another COBOL program which formats the data, producing six different reports. In the SPCC application, a COBOL program is used to edit and reformat the input data. A COBOL PLI program loads the changes into the data base. A COBOL PLI program will

produce a Spill Report under certain conditions as a result of the updates. Ad Hoc queries in System 2000 language produce output from the system.

Users: The majority of the Air Pollution and Emissions Testing System users will be field engineers using query language of the Immediate Access Module. These engineers will be given a two day training in the System 2000 query language. Maintenance of the data base will be done by programmer/analysts at EPA Headquarters. The users of the SPCC system are subject matter specialists who query the data base from regional offices. Maintenance of the data base is done by programmer/analysts.

Experiences with System 2000: MRI installed System 2000, but did not completely interface the system with TSO. MRI returned to complete the installation. A separate contract was awarded to MRI for initial assistance in data base definition. EPA has not experienced any software problems. When problems do occur, the vendor takes care of them in a short period of time. Vendor support is through a separate contract with MRI. This method of system support minimizes the time required for production of operational application systems. EPA feels that the System 2000 documentation is adequate.

Overall Impression: EPA feels that System 2000 provides what they were looking for: (1) a data base system operating on two computer systems; (2) flexibility in file maintenance; (3) an end-user oriented query language; and (4) very quickly developed new applications. System 2000 does not provide an explicit indication when file reorganization is required. System performance degradation has been experienced. System performance can be improved when the user invokes the REORGANIZE command which re-shuffles only the pointer tables. Although four data bases can be opened simultaneously with the PLI modules, no multi-data base capability exists at present in the Immediate Access Module. However, the present applications at EPA do not require multi-data base inquiry processing capability.

3.10 TOTAL Used at Bureau of Labor Statistics

Name of Agency: Department of Labor
Bureau of Labor Statistics (BLS)

DBMS Packages Used: TOTAL

When Installed: Purchased in December 1972 for approximately $30,0000.

Background: In 1971 BLS decided to redesign their statistical programs. Evaluation studies were conducted for selection of a generalized data base management system for the BLS

data processing needs. This study resulted in the purchase of TOTAL in December 1972. Testing of TOTAL was done in-house. The testing procedures verified that TOTAL was adequate for their needs and it was put into operational use in September 1973.

Applications: The Bureau of Labor Statistics gathers and receives vast amounts of data from State and local governments. The information is mostly survey data on socio-economic phenomena, and in general, comprises numbers which frequently are aggregated into time series. For example, one of the thirteen applications using TOTAL is the Federal/State Employment Hours Earning System (E&E). This data, received from various State and local governments on magnetic tape, punched card or paper forms, are entered into the data base in batch mode. The outputs from the data base are statistics in tabular form on employment and earnings.

Computer Environment: The TOTAL package is installed on the BLS's IBM 360/65 as well as the IBM 370/168, operating under OS with 20K bytes of core. Although TOTAL is marketed with ENVIRON (a teleprocessing interface), BLS does not have this option since they operate only in batch mode at the present time.

Operation and Use: The TOTAL system resides in the library and executes as a regular batch task under OS. The volume of batch updates for the E&E is on the order of 180,000 per month. Batch queries are less numerous and are usually performed for bulk data extraction. Extracted data are then used as input to programs for statistical calculations and cross tabulation. The data base is kept on disks.

Application Data Characteristics: The largest TOTAL data base at BLS is over 200 million bytes. It is partitioned into about ten different files, of which two are Master files ("Master" in TOTAL's terminology means the collection of data elements that are single entry, quickly searchable and contain the root segments of a hierarchical set of data. The other files are referred to as "Member" files or variable entry files.) The data are principally numeric and are mostly fixed length.

Application Program Characteristics: About thirty specific COBOL application programs have been built around TOTAL for the E&E System. Other functional programs perform editing and screening of data and interfacing to statistical routines (called SOUPAC), cross tabulation routines (called TPL), plus photo composition, graphical display and microfilm output programs. TOTAL does not have built-in data type checking on input, nor data conversion on output. Application programs are needed for these functions. Although TOTAL is also

marketed with SOCRATES (a language for a report/display writer), BLS does not have this capability. All output programs are custom application programs. PL/1 is the major procedural language used for application programs.

Users: The majority of users are statisticians and economists who have had training in data base concepts and are able to submit batch runs to extract data from the TOTAL data base. The updates are performed by PL/1 and COBOL programmers.

Experience with TOTAL: To install TOTAL under OS, the user prepares the JCL to allocate the required space for the data base files. BLS first establishes a data dictionary. The definition effort is comparable to writing a COBOL file definition section. A utility program, called FORMAT, formats the disk area according to TOTAL requirements and a user-written load program is needed to compile, link and load the data values. Once installed, the data base is very easy to maintain. At BLS, this effort takes one day a month.

Vendor support is reportedly good. The vendor provides a two-day training program as part of the purchase price. BLS found this adequate to install TOTAL and now operates the system on their own. Documentation provided by CINCOM is rated as adequate and improving. The vendor's marketing philosophy is to keep TOTAL small and simple, therefore very little enhancement and augmentation have been provided. TOTAL is as close to a "bug-free" program as BLS has experienced.

Overall Impression: BLS considers that they do not need self-contained data base management systems, but rather an access method to organize the data base and deal with the physical storage structure. Since the output presentation of their data is customized, and the data are never retrieved by single queries, BLS feels that a query language with report writing capability is not a key requirement for them. BLS survey data have multiple values with hierarchical attributes and therefore require a DBMS which supports a "network" file structure.

TOTAL was the sole software package in late 1972 that supported a network data structure on the IBM 360/370 series of computers. TOTAL requires small amounts of core (basic package needs only 13K bytes), is inexpensive ($30,000 to purchase) and is easy to install.

3.11 TOTAL Used at Social Security Administration

Name of Agency: Social Security Administration, HEW

DBMS Package Used: TOTAL

When Installed: Purchased in 1972.

Background: Social Security Administration wanted a system to be quickly installed and flexible enough to build their own applications. Also, the key requirements were storage efficiency and speed of access. They did not have a requirement for Boolean conditional searches. Retrievals are almost always a single record inspection. A study was conducted and three DBMS were considered: IMS, TOTAL and ADABAS. ADABAS was relatively new in 1972. The heavy emphasis on storage efficiency was the determining factor in the choice of TOTAL. (The storage expansion factor of TOTAL is said to be 15% to 20% over the raw data.)

Applications: There are three applications on TOTAL which operate under the Social Security Network called SSADARS (Social Security Administration Data Acquisition and Response System). The largest application is the Supplemental Security Income Program. This data base contains records for about 5.5 million welfare recipients. This data base is queried on-line (to resolve problem cases) in field offices in different parts of the United States via SSADARS. Another application is the case-control application. This data base contains records of all cases which require special attention. A third application is the Message Intercept Facility which has no user interface. It uses TOTAL to maintain a list of all the remote printers that are not functioning, so that messages to the "down" printers can be intercepted, stored and later forwarded.

Computer Environment: TOTAL runs on an IBM 370/168 under VS-2 at the Social Security Administration. 150K bytes of core plus buffers are required to operate the system. There are more than 750 output terminals in the field offices communicating with the central computer via the SSADARS network.

Operation and Use: For the Supplemental Security Income Program application, approximately 100,000 on-line queries are made each day by the claims representatives. The nature of the queries are always a single record retrieval keyed by the Social Security Number. Conditional relations combined with Boolean operators are not used. All updates are done in batch mode. For the five and one-half million records of the Supplementary Security Income Program, there are approximately 200,000 changes per month. There are several batch programs

written mainly in COBOL that are used for statistical gathering purposes.

Application Data Characteristics: There are approximately 30 files in the data base. The data are mostly coded and of fixed length. There are no large amounts of textual data. Each record consists of a fixed part and a variable repeating group part. For the Supplemental Security Income Program, the average record size is 450 characters.

Application Program Characteristics: COBOL and assembly language are used for application programming. For example, an in-house input/output handler was written to get data from the TOTAL data base and translate it into a fixed format. Social Security has built an in-house query language for the field offices. The query language uses a transaction number for invoking a predefined process, and almost always one specific record is retrieved and displayed in response to a query.

Users: The field offices are given a procedure manual and the transactional codes for various types of retrievals. Updating and data entry are done by programmer/analysts. At present, Social Security has a large application programming staff and has five data base specialists.

Experiences with TOTAL: Social Security has had no problems with the installation of TOTAL. A short period of time was required to make TOTAL operational. The vendor is very responsive to users but new enhancements to TOTAL are slow. There is a TOTAL users group which meets twice a year. Social Security feels that the documentation could be a little more complete.

Overall Impression: TOTAL was developed with small users (DOS environment) in mind. It is simple to use. However, large multi-file users may experience difficulty because of the lack of reentrant coding and the lack of capability for accessing multi-file data bases. All the data files are locked out when one is being accessed for structural maintenance purposes. Therefore, simultaneous processing of several data files is impossible. Reorganization of data is necessary when the system runs out of space; performance degrades because of the addition of new variable data records over a period of time. Single record retrieval is very efficient with TOTAL and flexible user language can be built to suit in-house users.

4.0 APPLICATION CONSIDERATIONS OF DBMS

It is almost impossible to find an off-the-shelf DBMS which does everything and is exactly tailored for an intended application. Systems are built with different objectives and hence have different capabilities for different application situations. In the design and implementation of a DBMS, design decisions such as file organization and search strategy tend to gear the system towards certain applications. Recognition of those design factors is critical in mapping needs to capabilities.

Before choosing a system managers must make a definitive analysis of their requirements specifically noting which features are mandatory and which are desirable. Besides weighing the technical trade-offs, subjective factors such as time and cost constraints, political considerations, prestige concerns, and long range policy must also be considered.

In this section, seven application considerations are identified based primarily on system design features peculiar to specific application needs. These considerations should serve to alleviate the problems associated with selection of a suitable system.

Numeric Data Versus Textual Data Oriented - Even though all of the DBMS support both numeric and textual data, some systems offer more capabilities in handling numeric data while others concentrate more on processing textual data. Numeric data capabilities can include a range of arithmetic manipulations and built-in statistical routines. Textual processing capabilities can include not only the support of variable length text data fields but provide content search capabilities, for example, vocabulary control, full text inversion, stemming, synonym matching, display of related terms, etc. A "numeric data processing" system can best be used in financial, accounting, or scientific data system applications, while a textual data processing system is typically used in legal or bibliographic applications. Among the systems surveyed, none is considered to be a "textual data processing" system. All of the six possess numeric processing capability.

Retrieval Oriented Versus Update Oriented - The retrieval algorithms employed by each DBMS consist of trade-off design decisions which favor either fast retrieval or fast updates. For those that favor fast retrieval, the technique used is either an inverted index or a chaining mechanism. However, both techniques created added overhead in storage space and require more update time to adjust pointers. The application which requires fast retrieval can best be used in a "static" data base environment. A "static" data base does not require constant up-to-the-minute changing. Weekly or even monthly updates could be tolerated. Examples of such applications are a data base of survey information and a grants management data base. Other applications need fast retrieval and constant changes to the data base. Examples

of such applications are airline reservation systems and, to some extent, inventory control applications where most of the activities are updating of the data bases. Among the systems surveyed, ADABAS, INQUIRE, MODEL 204, and SYSTEM 2000 are particularly suited for quick retrieval while IMS and TOTAL are more oriented toward updating activities.

Self-Contained Versus Host Language - The "self-contained" DBMS provide the users with a language for updating, quering and report writing while those generally referred to as "host-language" DMBS require writing application programs in a procedural language with embedded calls for data base functions. The self-contained systems in this study are INQUIRE, MODEL 204, and SYSTEM 2000. All of the systems surveyed have host-language interfaces and thus can be considered host-language systems, but IMS (stand-alone) and TOTAL can be used only as host-language systems. ADABAS started as a host-language system, but has recently provided a query language ADASCRIPT and a report writer ADAWRITER. Thus, ADABAS is evolving from a host-language to a self-contained system. IBM has provided an optional query language (Interactive Query Facility) and has recently announced the marriage of GIS (Generalized Information System) with IMS. Thus, the couple is a self-contained system.

Both types of systems have advantages and disadvantages. The self-contained types are end-user oriented with little or no application programming necessary. However, the user is paying for an added layer of software in terms of efficiency and flexibility. On the other hand, the host-language type systems provide all the necessary data management functions plus the total power of a programming language for manipulating data.

Ad-Hoc Query Versus Pre-Determined Transaction Processing - Another application paradox is whether the queries of the data base are predetermined or whether they are unforeseeable. Unanticipated queries or ad-hoc query requirements usually have to pay the penalty of a sequential scan if the needed data elements are not keyed or inverted. For example, in a payroll or financial data system application, the user could essentially be allowed to formulate queries with any data elements in any Boolean combinations. On the other hand, "transactions" are normally pre-defined functions such as daily output reports with known sets of parameters. Most applications need both capabilities but the ratio and relative importance of ad-hoc queries to pre-determined transactions must be defined before system selection.

Pre-Defined Process Via User Language Macro Versus Procedural Programming Language - Transaction-oriented functions such as daily reports or new data record entries which are used over and over again, are customarily pre-constructed and stored in the system. These pre-constructed functions called "pre-defined processes" can be created either by writing a subroutine in a procedural programming language such as in COBOL, or by using a macro facility with parameter substitu-

tion capabilities provided by the DBMS. Such macros are pre-constructed by using the DBMS provided language. It is generally contended that DBMS provided user languages are higher-level than programming languages and therefore require less skill and time than writing program procedural language such as in COBOL. However, they may be less flexible and less efficient in programming. INQUIRE, MODEL 204 and SYSTEM 2000 all provide user language macros, while ADABAS, IMS/VS and TOTAL do not.

Network Versus Non-Network Structures - DBMS usually provides some means for expressing relationships among data elements. The types of data structures supported within a DBMS imply certain applications by the way in which data may be related semantically. For certain applications, data relationships are hierarchical, e.g., parent-child relationship, while in other cases, more than uni-directional hierarchical relationship is required. Such relationships are generally referred to as network relationships. The principal disadvantage of network data relationship is the complexity of data maintenance. For example, a "delete" operation in a hierarchical relationship deletes the parent together with all its descendants, while a "delete" operation in a network relationship requires checking to make sure no other data are related to deleted elements. TOTAL is the only system in this study which supports network relationships.

Large "Public" Versus Small "Private" Data Base - The problems associated with processing a data base of all motor-vehicles licenses in the country, or of social security recipients are very different from processing a private list of items such as equipment, books on order, etc. In the first place, the volume of data that needs to be maintained is different, and secondly, the workload of the application is much greater. It is important to realize that certain DBMS simply cannot take heavy workload with large data base sizes without degrading performance, while some DBMS were built with large users in mind. Unfortunately, no DBMS vendors will admit that their DBMS cannot handle "big" users. Projected performance criteria cannot be assessed unless the DBMS has been subject to benchmark tests.

Figure 4 is a matrix summarizing the above application criteria. This matrix does not include performance considerations of these systems with different size data bases, but the user's experiences provide insight into how each system performs with respect to its data base size and the number of terminals operating simultaneously.

An overall evaluation of DBMS must include these application criteria together with considerations of computer configuration, operational environment, and growth potential.

	ADABAS	IMS/VS	IMS/VS with GIS	INQUIRE	MODEL 204	SYSTEM 2000	TOTAL
Numeric Oriented	X	X	X	X	X	X	X
Textual Oriented							
Retrieval Oriented	X		X	X	X	X	
Update Oriented		X					X
Self-Contained Query	X		X	X	X	X	
Update	X		X	X	X	X	
Report Writing	X		X	X	X	X	
Host Language Interface Query	X	X	X	X	X	X	X
Update	X	X	X	X	X	X	X
Report Writing	X	X	X	X	X	X	X
Ad-hoc query	X		X	X	X	X	
Pre-determined transaction		X	X	X	X	X	X
Pre-defined Process via User Language Macro			X	X	X	X	
Pre-defined Process via Procedural Programming Language	X	X	X	X	X	X	X
Network Data Structure							X
Non-Network Data Structure	X	X	X	X	X	X	

Figure 4. Summary Matrix of DBMS Application Criteria

REFERENCES

[1] Schussel, G., "Data Base Systems Wave of the '70s," Computerworld Special Report, February 27, 1974, page S/2.

[2] Datapro Research Corporation, "A Buyer's Guide to Data Base Management System," Datapro feature report, Delran, N.J., November 1974.

[3] CODASYL Systems Committee, "A Survey of Generalized Data Management Systems," CODASYL Systems Committee Technical Report, May 1969.

[4] CODASYL Systems Committee, "Feature Analysis of Generalized Data Base Management Systems," CODASYL System Committee Technical Report, May 1971.

[5] Anderson, R. K., "Index of Data Management Software Packages," National Bureau of Standards, Report 10 932, October 1972.

[6] Fong, E., "A Survey of Selected Document Processing Systems," National Bureau of Standards Technical Note 599, October 1971.

[7] Fife, D.W., et al., "A Technical Index of Interactive Information Systems," National Bureau of Standards Technical Note 819, March 1974.

[8] Marron, B., et al., "A Study of Six University-Based Information Systems," National Bureau of Standards Technical Note 781, June 1973.

[9] Fry, J. P., et al., "Data Management Systems Survey," MITRE Corporation Report MTP-329, January 1969.

[10] Koehr, G. J., et al., "Data Management Systems Catalog," MITRE Corporation Report MTP-139, January 1973.

[11] Martin, T. H., "A Feature Analysis of Interactive Retrieval Systems," Stanford University Report SU-COMM-ICR-74-1, September 1974, (NTIS No. PB-235952/AS).

NBS-114A (REV. 7-73)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBS TN-887	2. Gov't Accession No.	3. Recipient's Accession No.

4. TITLE AND SUBTITLE

SIX DATA BASE MANAGEMENT SYSTEMS: FEATURE ANALYSIS & USER EXPERIENCES

5. Publication Date: November 1975

6. Performing Organization Code

7. AUTHOR(S)
Elizabeth Fong, Joseph Collica, Beatrice Marron

8. Performing Organ. Report No.

9. PERFORMING ORGANIZATION NAME AND ADDRESS

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

10. Project/Task/Work Unit No. 640-1414

11. Contract/Grant No.

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

Federal Aviation Administration
800 Independence Avenue, S.W.
Washington, D. C. 20591

13. Type of Report & Period Covered: Final

14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

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

This report presents an objective overview of features of six selected data base management systems (DBMS) and Federal user experiences with these systems. Application criteria were developed in order to aid in the evaluation and selection of DBMS.

The advantages of generalized DBMS over traditional methods of software system development are discussed. The criteria for choice of the six systems' features are presented. The data reported were gathered from two sources: vendors and users. Vendor information consisted of side-by-side presentation of features of the six DBMS. User experiences reported are summarized under appropriate headings. This information is used to derive application criteria for assessing the usability and operational suitability of DBMS to a variety of data processing requirements.

The six systems reviewed are: ADABAS, IMS/VS, INQUIRE, MODEL 204, SYSTEM 2000, and TOTAL.

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

Data base applications; data base management systems; selection criteria; software features; system evaluation; system features; user experiences.

18. AVAILABILITY

[X] Unlimited

[ ] For Official Distribution. Do Not Release to NTIS

[X] Order From Sup. of Doc., U.S. Government Printing Office Washington, D.C. 20402, SD Cat. No. C13.46:887

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

19. SECURITY CLASS (THIS REPORT): UNCLASSIFIED

21. NO. OF PAGES: 84

20. SECURITY CLASS (THIS PAGE): UNCLASSIFIED

22. Price: $1.45

USCOMM-DC 29042-P74

NBS TECHNICAL PUBLICATIONS

PERIODICALS

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

- **Physics and Chemistry (Section A)**

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

- **Mathematical Sciences (Section B)**

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

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

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

NONPERIODICALS

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

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

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

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

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a world-wide program coordinated by NBS. Program under authori of National Standard Data Act (Public Law 90-39(

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

Building Science Series—Disseminates technical info mation developed at the Bureau on building material components, systems, and whole structures. The serie presents research results, test methods, and perform ance criteria related to the structural and environmen tal functions and the durability and safety character istics of building elements and systems.

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

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

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

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

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

Order NBS publications (except NBSIR's and Bibliographic Subscription Services) from: Superintendent of Documents, Government Printing Office, Washington, D.C. 20402.

BIBLIOGRAPHIC SUBSCRIPTION SERVICES

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

Cryogenic Data Center Current Awareness Service

A literature survey issued biweekly. Annual subscription: Domestic, $20.00; foreign, $25.00.

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

Superconducting Devices and Materials. A literature survey issued quarterly. Annual subscription: $20.00. Send subscription orders and remittances for the preceding bibliographic services to National Technical Information Service, Springfield, Va. 22161.

Electromagnetic Metrology Current Awareness Service Issued monthly. Annual subscription: $24.00. Send subscription order and remittance to Electromagnetics Division, National Bureau of Standards, Boulder, Colo. 80302.

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

OFFICIAL BUSINESS

Penalty for Private Use, $300

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

SPECIAL FOURTH-CLASS RATE
BOOK