PNAAQ471
ISN 35525
263016
62

DATABASE SOFTWARE SELECTION FOR THE EGYPTIAN NATIONAL STI NETWORK

Vladimir Slamecka

USAID Contract NEB-0016-C-00-2092-00

June 1984

Georgia Institute of Technology
Atlanta, Georgia

Academy of Scientific Research and Technology
Cairo, Egypt

ABSTRACT

The study reports on the evaluation and selection of information/data management software for the Egyptian National Scientific and Technical (STI) Network. The network consists of sectorally oriented information service nodes (ISNs), each having a responsibility for the development and maintenance of public databases of Egyptian information and data.

The report elaborates on the differences between information retrieval and database management (DBMS) systems, and concludes that while the Egyptian databases are likely to contain both unstructured text and structured numeric data, the former are likely to receive initial emphasis. From this premise the report lists the mandatory and desirable specifications of the database software.

The evaluation methodology was comprised of an extensive, two-year survey of available commercial software and its evaluation against the mandatory specifications. The strategy of selecting a software package included, in the order of preference, a) a "universal" DBMS capable of handling both text and data types; b) a relational DBMS expandable to handling full-text retrieval; and c) an IR software system. After studying nearly 100 software packages it was concluded that no "universal" DBMS and no relational DBMS on the market meet the constraints and demands of the National STI Network of Egypt.

The recommended software is the recently introduced mini/micro version of BRS/SEARCH, a full-text information management software package from BRS, Inc. The GSA software evaluation form and other documentation regarding the recommended package are included in the report.

CONTENTS

ABSTRACT 3

A. INTRODUCTION 7

B. TYPES OF INFORMATION SYSTEMS 9

1. Information Retrieval Systems 9
2. Database Management Systems 10
3. Distinctions Between IR and DBMS Systems 12
4. Conclusions 18

C. EGYPTIAN STI NETWORK SOFTWARE SCHEMA 21

1. Database Building and Management 21
2. Database Search and Retrieval 23
3. File Creation and Processing 23
4. Communications 23
5. System Administration 24

D. EGYPTIAN STI NETWORK DATABASES 26

1. Development of Sectoral Database Pools 26
2. Contents of Sectoral Databases 29
3. Nature of Data in Databases 34

E. EGYPTIAN STI NETWORK DATABASE SOFTWARE REQUIREMENTS 35

1. National-Standard Features 35
2. End-User Orientation 36
3. Performance 37
4. Security 39
5. Modifiability 40
6. Support 41
7. Database Definition 42
8. Database Input and Maintenance 42
9. Database Search and Processing 43
10. Search Output and Processing 44
11. Administration 44

F. SURVEY OF COMMERCIAL IR/DBMS SOFTWARE 45

1. Database Software Survey 45
2. Previous Comparative Studies 46
3. Preliminary Review 46
4. Analysis of Candidate Software 46

G. FINDINGS 49

1. "Universal" Database Software 49
2. Relational DBMS Software 51
3. Information Retrieval Software 54

H. RECOMMENDATIONS 56

1. Information Management Software 56
2. Data Management Software 57
3. Administrative DBMS Software 58

I. REFERENCES 59

APPENDICES
I: Status of IV+V Software 67
II: Evaluation Form 71
III: Interviews 84

LIST OF FIGURES

1. Architecture of Information Retrieval System 16
2. Overall Schema of Software Facilities 22
3. ISN Databases: Input 27
4. ISN Databases: Output 30
5. A Generic Problem Solver Paradigm 31

LIST OF TABLES

1. Dominant Characteristics of Information System Types 20
2. U.N. Organizations Operating Systems 28
3. Factual Data: The Energy Sector 33
4. List of Reviewed Software 47
5. IR/DBMS Software 50
6. Relational DBMS Software 52
7. Information Retrieval Software 54

A. INTRODUCTION

The purpose of the Egyptian National Scientific and Technical Information Network is to provide Egyptian problem solvers access to public data and information relevant to their professions and to the social and economic development of the country.

The STI Network is organized along the sectoral dimension of the national economy, so as to place the management of public information resources and services close to the domains of technical and professional expertise and need. In each sector there is to exist a designated information agency functioning as the lead sectoral "information service node" (ISN). The principal responsibilities of the sectoral ISNs are 1) to facilitate access by its sectoral clientele to repositories of problem solving data and information that reside at the node itself, elsewhere in Egypt, or abroad; and 2) to organize or procure, maintain, and make publicly available machine-readable repositories of data and information at the ISN. These repositories, or "databases," available for public access at each ISN may come from several sources: 1) building of Egyptian databases by the ISN; 2) lease of databases, or segments thereof, developed by other Egyptian organizations, including other sectoral ISNs; and 3) lease of foreign databases.

The STI Network and its information service nodes are thus considered vendors, and in some cases copyright owners, of publicly available, machine-readable data and information. In providing information services to the public, the STI Network nodes do not function as computer centers, data centers, or management information systems for any organization, including their own parent agency (so called "host organization"). The objective is to collect and provide public-domain data and information; classified information is excluded from the purview of the STI Network.

Database building activities and information services are coordinated at the STI Network management level, with participation of the sectoral STI nodes. Equitability among the nodes is assured via Network-wide principles governing such aspects as fees for information services, copyright and royalty arrangements for distribution of databases, and planned development and expansion of the STI Network. While being conducive to equitability, these principles are not intended to stifle healthy competitiveness among the members of the STI Network.

Networking also implies voluntary conformity to certain agreements of compatibility and to common standards. Among these, the most important is the principle of Network transparency to its users. Whereas the data and information resident in different sectoral nodes of the STI Network will vary considerably, as will the use of this data and information by problem solvers, the means (called "protocols") of access and interaction with the databases should be as uniform as possible. For this reason, the STI

Network strongly endorses the principle of portability of both information/data resources and procedures of using them.

At the level of technology, this principle is supported at the software rather than hardware level, via a portable operating system that allows hardware of divergent manufacturers to operate in a network environment. At the procedural level, transparency of system facilities from the users' viewpoint is sought via a robust set of uniform data- and information-management software facilities that would enable, on the one hand, the building and maintenance of relatively uniform information and data repositories, and on the other hand would offer a similarly uniform set of capabilities and protocols for database searching and output manipulation. These software facilities are referred to popularly as "database management systems," or DBMS software.

This document concerns the characteristics and selection of the DBMS software for the Egyptian National STI Network. Following an overview of the state of the art of database technology, the document identifies desirable characteristics of such software for Egypt, reviews existing commercial DBMS systems, and recommends an appropriate strategy for the Egyptian National STI Network.

B. TYPES OF INFORMATION SYSTEMS

In the still evolving discipline of information science and technology, the term "information system" does not have a precise meaning; rather, it covers a range of entities, the delineation among which is also not very precise. Paraphrasing and modifying the categories described in Salton and MacGill [78], we may make useful distinctions between two principal classes of information systems: 1) information retrieval systems, and 2) database maragement systems. Each class will further exhibit several typical variants of its genre. Their overall characteristics may be summarized as follows.

1. Information Retrieval Systems

The traditional context in which the phrase "information retrieval systems" is best known is that of the representation, storage, and access to documents or document surrogates. The input to such systems is likely to include the natural language text of documents or document excerpts (abstracts, summaries). The output of online or batched searches consists of sets of references to documents called "citations," or (in the case of documents stored in full-text) of portions of document texts. The output is generated in response to a "search query," the form of which usually incorporates a simple syntax and natural language "keywords" or short phrases.

Although the document itself tends to consist predominantly of natural language text, and hence of relatively unstructured data, some of its elements may be highly structured (e.g., tables). Except when the document is entered into the informations system in toto, its representation is given in a relatively structured form.

Documents are not the only content of information retrieval systems, however; the term "document" may be extended to include any discrete statement in natural language. Falling within the range of such a definition are memoranda, research notebooks, survey forms, newspapers and newsletters, organizational plans, maps, recordings, art, museum objects, hospital patient records, legal contracts, and other entities. Depending on the contents of the files, such systems are called museum information systems, and the like. The users of information retrieval systems have widely different backgrounds and a wide variety of different information needs; the largest traditional category of users have been researchers, R&D managers, and members of diverse professional communities.

2. Database Management Systems

A database is a collection of stored items; and a database management system is a facility for manipulating and maintaining control of the database. In practice, database management systems facilitate the storage and maintenance as well as retrieval of data from the database.

The traditional notion of database management systems tends to view the contents of the database as "data" rather than "information"; that is to say, they consider the content of a record field to be the value of that field, devoid of semantic content. The value may exist as a string of numeric, alphabetic, or alphanumeric characters, and it may even be a natural language string; in this case, the tendency, however, is to consider that string a datum -- e.g., a name defining a person, an object, etc.

Database management systems facilitate the storage and retrieval of data, its updating or deletion, and protect the data from unintentional or deliberate misuse. Search requests of a database must state the specific values of certain record identifiers; the retrieved information consists of all records which match the stated request exactly. This last characteristic of database management systems is often said to be their key difference from information retrieval (see below).

Database management systems tend to include more numeric data than do information retrieval systems, and they are frequently supplemented with computational or statistical facilities that operate on the database itself. Depending on the objective of such computations and data manipulations and displays, several discrete categories of DBMS-based information systems have evolved, as summarized below. All of these categories store the data in a "database," and all employ some of the basic software facilities of DBMS systems (for data input, modification, retrieval, etc.). They are distinguished from each other primarily by the additional data-manipulating functions that are added on top of the DBMS package and that perform the discrete objectives of the information system category.

a) Management Information Systems. The customary perception of an MIS is that of a database management system tailored to the needs of organizational or corporate management. While fitting the general database management image, an MIS typically contains facilities of interest to managerial functions (data aggregation, modeling), and the database tends to contain appropriate management-oriented data (financial, personnel, inventory, production). Management information systems are thus a special purpose DBMS category.

Whereas the concept of management information systems is not new, it has recently evolved to a more generic concept called "enterprise information systems" [56]. The latter provide a data-oriented view of processes extant in an organization, and database management systems are a principal tool that supports this view.

b) Decision Support Systems. DSS exemplify a still further extension of management information systems. Whereas MIS perform specific operations on homogeneous classes of data or information, decision support systems assemble a variety of facilities into a single structure for the purpose of supporting decision-making processes. Currently such facilities include database management, information retrieval, computer graphics, financial and other data analysis and modeling, and sophisticated report generation. The gradual integration of progressively more complex facilities into decision support systems is the current trend in office systems development; as such systems increase in their generality, they become of interest to broader ranges of users at both the strategic and non-strategic levels of decision-making [29,51].

c) Question-answering Systems. Question-answering systems currently exist principally in experimental stages [27]. Their purpose is to provide access to factual information in a natural-language setting: natural language is used to submit a query to the system, and the latter also responds in natural language.*

d) Expert Systems. This new category of information systems encompasses a database management system, in which the database content consists of decision rules employed by a human expert within a specific, usually narrow domain of expertise. Expert systems are intended to emulate the decision-making of human experts in various domains, so as to assist less experienced problem solvers. The decision rules are of the "if . . . then" variety. Expert systems are interactive, and are capable of eliciting from the user enough relevant data to invoke an appropriate decision rule; their ability to then explain why the particular rule has been invoked is another characteristic feature.

The database management component of expert systems is responsible for housekeeping of the decision rules, and it provides a relatively sophisticated, natural-language, conversational interface between the user and the database. Expert systems exist for an increasingly larger number of narrow domains (e.g., infectious skin diseases, repair of electronic devices) and they are a cornerstone of the emerging disciplines of "knowledge engineering" and artificial intelligence.

---

*Question-answering systems are different from so-called natural language query systems. The latter refer to the ability to phrase questions to a database using the regular syntax of natural language, and thus fall into the category of "query languages," or user interface facilities. In the past two years a small number of natural language query facilities have become available and are gradually being incorporated into database management systems [21, 100].

3. Distinctions Between Information Retrieval and DBMS Systems

Before proceeding to analyze the functional requirements of the database software for the Egyptian National STI Network it is helpful to clarify a number of distinctions that characterize the functions of information retrieval and data retrieval. We review here the following classes of such distinctions: 1) user objectives and expectations, 2) query languages, 3) data architectures, and 4) flexibility.

a) User Objectives and Expectations. Major distinctions between information retrieval and database management systems exist with respect to the goals and expectations of end users.

In data retrieval from the DBMS, a query retrieves the actual data required (for example, the salary of employee X). A document retrieval system is more indirect; it retrieves not the particular datum requested, but instead provides pointers to one or more documents that are likely to contain the desired information. In information retrieval, as opposed to database management systems, requests are not formulated precisely but are posed more generally than data queries; and the retrieved information may include items that do not match the request. However specific a document retrieval system is, it is never as specific as a typical data retrieval system.

Information retrieval thus has a probabilistic aspect to it; this is to say that, from the logical viewpoint, such systems are essentially non-deterministic. Document retrieval is fundamentally a decision under uncertainty. In contrast, data retrieval from a DBMS is deterministic, in that the values of the query statement are matched exactly against the values contained in the record fields and elements, with no semantic uncertainty being involved. In terms of decision theory, data retrieval represents a pure reference decision.

Since results from information retrieval queries almost invariably contain varying degrees of irrelevant material, much effort over the past 20 years has gone into research to reduce this "imprecision" of information retrieval systems (the characteristic that they retrieve less than all the relevant items from the database, while also retrieving varying amounts of irrelevant items) [8,21,30,31,45,54,55,68,70,76,98,99]. By and large, however, the results of this research have not been incorporated in commercial information retrieval systems, for a variety of reasons. Consequently, information retrieval software has not made many meaningful strides toward approaching the retrieval precision of data retrieval systems.

Differences between the purposes of information retrieval and DBMS systems are also reflected in the query process itself. DBMS queries are formulated typically as a single statement, and their output constitutes

the answer. In information retrieval the output is very often unsatisfactory (too large, too small, irrelevant, etc.), and may require further manipulation (limiting it to a subset, expanding, etc.). This distinction affects the query languages of the two types of information systems (see below).

The different expectations of users of information retrieval vs DBMS systems are also reflected in their respective notions of system "goodness." For the information retrieval system user, the primary criterion of system performance is his/her satisfaction: if system A provides a more useful set of documents than system B, it has greater utility to the user. For data retrieval from a DBMS, the criterion of success is correctness of the data retrieved, as the user expects to either retrieve the exact data sought, or be advised that the database does not contain it.

As another system performance criterion, retrieval speed is of considerable importance to a DBMS user; he is far more conscious of it than is the user of an information retrieval system. The latter rarely retrieves all and only the desired information, and expects to dedicate considerable effort to review the query result by sifting the useful references from the irrelevant ones. In comparison with this effort, even a slower-response system often appears adequate.

b) Query Languages. The set of commands used to negotiate a search in a database is of primary concern in both classes of information systems, in that it is the primary factor affecting the "friendliness" and ease of the interaction. User-friendly commands are implemented via so-called high-level query languages; they are used to state what the user wishes to accomplish (in contrast to low-level languages which require the user to specify how the goal is to be accomplished). Extensive reviews of query languages exist in the literature [3,61,83,90,91].

Ideally, it might be useful to provide the unrestricted natural language for the formulation of database queries. Although still in the nascent stage, software is now on the market that allows users to ask database queries in natural language. The complexity and the size of the programs of such software are substantial, and hence the software is initially running on large database management systems; at least in one case, such software does run on a microcomputer DBMS, however. At this point no information retrieval system allows the use of queries to be formulated in unrestricted syntax with unrestricted vocabulary; some such systems do, however, permit the use of restricted natural language syntax.

To accommodate the occasional user who is typically unable to remember large numbers of commands and details of precise syntax, several types of query language facilities exist. In the order of increasing sophistication, they include the following:

i) Responding "yes/no" to questions displayed on the terminal;

ii) Selecting one item or action from a list ("menu") of multiple

choices displayed on the terminal;

iii) Filling a blank space on the terminal with an appropriate keyword or command in response to a "prompt" by the computer;

iv) Filling in a table with more complex linguistic and/or numeric structures (e.g., keywords in a set theoretic, and/or values related in a numeric way);

v) Using restricted natural language syntax to specify a set of operations on database fields (e.g., "select x,y from database D where x = abc and y > p");

vi) Using natural language with less restricted syntax.

Information retrieval systems currently on the market employ predominantly approaches iii) and iv) for keyword-based queries; understandably, because of the large number of potential keywords, neither i) nor ii) is applicable. Whenever an operation can be selected from a small number of alternatives (not larger than one frame or screen), approaches i) and ii) are preferred; for instance, in the selection of one from several databases. Relational DBMS queries employ approaches v) and, in a smaller number of instances, iv). In principle, there is no reason why relational DBMS software cannot use the fill-in-the-blank technique. The typical software of an information system is likely to use more than one technique for user-system interface.

Standardization of query languages leaves much to desire [3], as does database standardization in general [17,97]; progress is being made, however, with respect to query languages for bibliographic information retrieval and relational DBMS systems. While no international standard exists as yet, the Common Command Language (CCL) is an attempt of the European Economic Community (ECC) to reach consensus regarding the minimal sets of user commands for bibliographic information retrieval [61]. Similarly, the underlying theoretical foundation of relational DBMS provides a de facto standard syntax and functions to construct database queries [11,17]; different software packages will, nevertheless, use different commands and mnemonics for these functions.

c) Data Architectures. In the context of the present study, data architecture refers to the conceptual model for organizing the contents of information systems -- the data and their relationships. Extensive treatment is given in numerous textbooks [9,34,72,78,101]; the following observations are intended for the more casual, pragmatic reader.

The data architecture of information retrieval systems is based on designs that date to the 1950's. Document records are stored sequentially in a file (called "ordered sequential file"). To retrieve a particular record, two methods are employed: a) a sequential search, in which each user-specified search "descriptor" (keyword) is matched against the whole file, sequentially from beginning to end; or b) the inverted index search, in which such descriptors are matched against entries in a separate index

to the file (in an inverted index of keywords, each entry becomes a key to the location of corresponding records in the file). A typical architecture of an information retrieval system is shown in Figure 1.

Each search technique has its advantages and drawbacks. Compared to inverted index searching, sequential file searching is inefficient when files are large, even when using techniques such as binary search. On the other hand, the need to update the index every time a record is added or deleted from a file requires considerable processing effort, particularly with full-text files that are indexed on almost every alphanumeric string. In general, the background processing needed to generate an inverted index has been considered less objectionable than excessive retrieval times, and virtually all information retrieval systems of significant size employ the inverted-index search technique. In the last year, however, advances in hardware array architectures have led to the development of at least one commerical system (by General Electric Corp.) that employs sequential file searching, and claims retrieval speeds 100 times faster than the inverted index search. The initial cost of this hardware (on the order of $250,000) is mitigated by the advantage that no indices need to be generated and maintained.

As information retrieval systems begin to incorporate full text, the trend is likely to shift toward constructing specialized hardware architectures for text retrieval [33]. For the time being, the sequential-file and inverted-index data architecture is commonplace for document-oriented information systems, particularly those storing full text.

In contrast to the sequential file-oriented data architectures of most document retrieval systems, database management systems have, over the years, developed advanced system architectures that are traditionally classified as falling into three types of "models": hierarchical, network, and relational. For our purposes, the models differ primarily in the way they handle relations between record elements as well as between records.

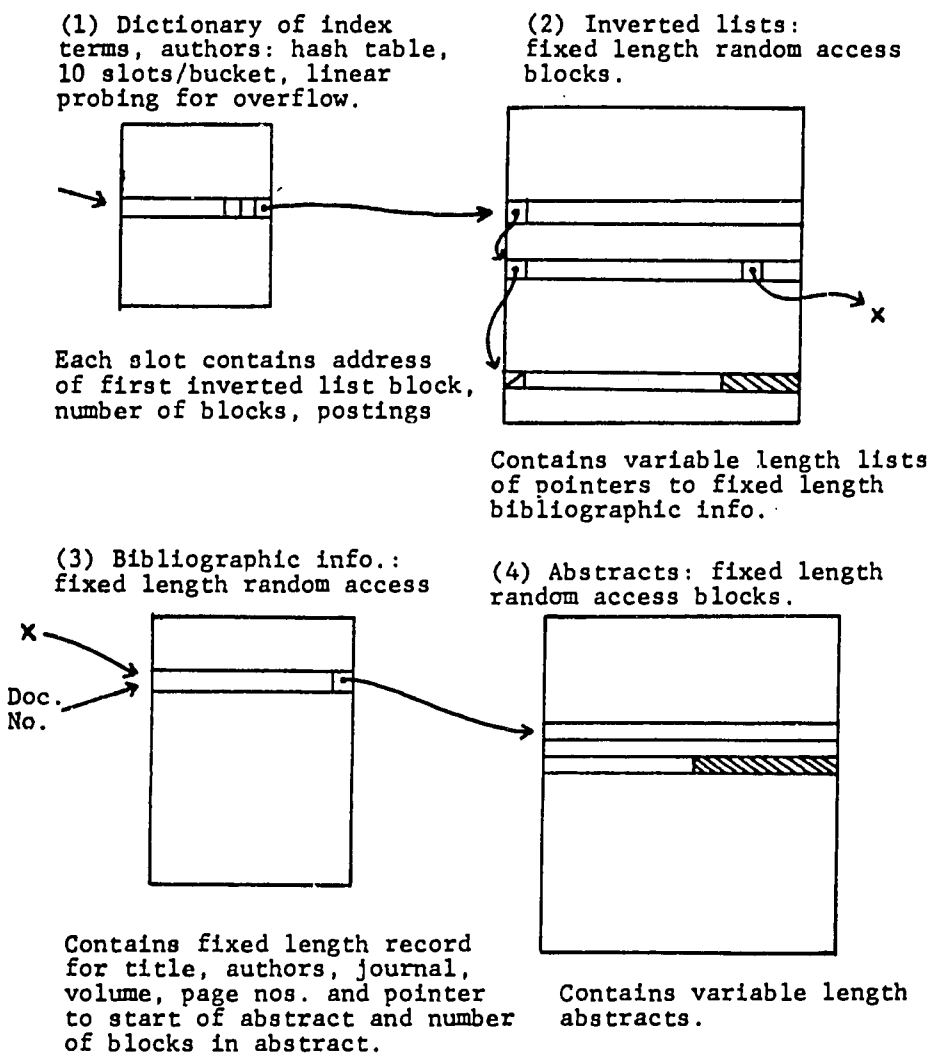

Figure 1. Architecture of Information Retrieval System (from [15]).

Hierarchical and network models date to the early 1960's. Both structure data rigidly into pre-defined relationships at the time the database is created; in querying the database, therefore, the user must "navigate" through the predetermined structure. Because the relationships are stored as chain pointers and the data items that form the chains have no internal relationship, they typically remained isolated from the user's query; the network structure is therefore more difficult to use. Any record occurrence takes on its full significance only when seen in context: dependent records cannot exist without superior records. Consequently, both users and programmers experience hardships with the hierarchical and network approaches. Users must be knowledgeable about the database and data structures, in order to navigate successfully through them. Programmers, on the other hand, typically require 9 to 12 months for learning and understanding a database management system based on the hierarchical or network approaches, and have to cope with considerable complexity of programs.

The relational approach owes its formulation to Codd, in the early 1970's [11]. It offers a logical view of data: all information (both entities and relationships) are presented in a uniform way as tables or "relations." The files are thought of as simple two-dimensional tables; there are no pointers out of the file to other stacks of data. Because of this, only one operator is needed for each of the basic functions such as "insert" or "delete." As a result, the relational structure is easy to understand and manipulate, and the number of distinct operations in any given language is small because there is only one type of data construct with which to deal. A typical programmer learning curve for a relational database is three months.

The easy conceptualization of the relational model by both programmers and end users has lent the relational approach a substantial advantage, to the point where DBMS software developed in the past three years for small and medium sized databases is almost invariably relational. Until recently there has been a paucity of experience with the performance of relational database management systems for very large data files. Some claim that relational databases tend to be slower than network or hierarchical databases, and hence less suitable in applications which require nearly instantaneous retrieval, such as airline reservation systems, military systems, and the like; careful comparative evaluations have, however, been few [67]. For information retrieval systems, however, speed of retrieval is less critical.

d) Flexibility. An important difference between information retrieval and some DBMS software has to do with its ability to be adapted to different user environments and functional needs. Typically, information retrieval systems are cast in concrete: they provide a fixed set of functional capabilities (which may include options) that is almost impossible to tailor to idiosyncratic needs of different users and user sites. The reason for this inflexibility is partly historical: the designers of these systems did not believe that there would ever occur a need for additional capabilities. As a result, bibliographic retrieval

systems do not, for example, support the simultaneous searching of two or more databases.

Hierarchical and network database management systems can be tailored to local requirements but only with a substantial effort by highly trained database programmers. As noted earlier, the reason for this is the complexity of the software and the need to be thoroughly familiar with data structures.

Current developments in computer science, especially in regard to fourth generation languages, are rapidly changing this situation. Relational DBMS software developed over the past three years provides database managers with built-in tools allowing them to modify database features to suit particular clienteles. As a result, one can conceive and implement, within the same database system, more than one query language, such that both casual and skilled users can be served efficiently. This set of tools (sometimes marketed independently as a so-called applications generator) contains facilities (actually languages) for creation of screens and forms, generation of reports, modification of query languages, and the development of other utilities. Application generators speed up the development or modification of database functions by an order of magnitude, and their use does not require extensive training.

Another technique by which present-day relational management systems add flexibility is by providing easy interfacing with the utilities of operating systems. When such utilities can be used with little training (as is the case of the UNIX shell), they provide a powerful way of generating new or modified applications and friendlier user interfaces.

The need to provide flexible but easy to use facilities for local tailoring of software is now recognized, and we should expect new information systems software, for both data types, to include them.

4. Conclusions

Table 1 is an attempt to present the dominant characteristics of information system classes as regards their contents, structure, query languages, and functions. The table confirms that there exist at present differences in objectives and functions between the two main classes of information systems -- information retrieval and database management software.

The question that is of some theoretical and considerable practical import is the following: is it possible to conceive of and develop "universal" software capable of accommodating all data types and all user categories? There have been a few attempts to do so (e.g., IBM's STAIRS software which features, however, separate query languages for information retrieval and DBMS operations). The question has received some attention in the research literature [13,19,50,78,88], although the final word regarding its feasibility may not have been said. The way to go appears to be to endow a relational database management system with capabilities to

handle unrestricted strings of unformatted text data. State-of-the-art commercial relational DBMS software is beginning to move in that direction, although slowly. Relational database software designed with information retrieval in mind is extremely rare (MINISIS is possibly the only such system currently operational). The converse alternative -- to adapt information retrieval software for other data types -- appears suboptimal because its file oriented data architecture is less attractive for data handling and retrieval.

As will be indicated later in this document, there are several efforts underway to design relational DBMS software capable of accommodating unformatted, free text data. Unfortunately, none of these efforts are likely to reach a commercial stage by fall of 1984.

Table 1.
Dominant Characteristics of Information System Types

	IR	DBMS	DB MIS	BASED DDS	EXPERT	Q/A
CONTENTS						
Alphanumeric Text (as information)	●	O		O		●
Alphanumeric Text (as data)	O	●	●	O	●	O
Numeric (as value)	O	●	●	●	O	O
STRUCTURE						
File	●					●
Hierarchical	O	O	O	O		O
Network	O	O	O	O		O
Relational	O	●	●	O	●	
QUERY LANGUAGE						
Natural Language (full syntax)		O	O			●
Natural Language (restricted syntax)	O	●	●	O	●	O
Keyword (Query-by-Example)		O	O			
Menu	O	O	O	O	O	
Keyword (Fill-in-blanks)	●	O	O	O	O	
FUNCTIONS						
Retrieval (determ.)	O	●	●	O	O	●
Retrieval (prob.)	●			O	O	O
Statistical processing		O	O	O		
Decision modeling		O	O	●		
Graphics		O	O	●		
Expert advice					●	

● = primary characteristic
O = secondary characteristic

C. EGYPTIAN STI NETWORK SOFTWARE SCHEMA

Figure 2 presents the overall schema of software facilities recommended for the Egyptian National STI Network. The functions shown are to be available and implemented at each sectoral information service node (ISN).

Broadly speaking, users of the computational facilities at each ISN fall into two categories: node staff (employees of the information service agency), and "external" users. The latter will vary considerably in respect to their computational skills; but it should be assumed that their majority will fall into the category of "casual users." Since by definition the STI Network is intended to serve the public, its facilities must initially be geared to accommodate the casual user.

As Figure 2 shows, all users, whether inhouse or external, and whether Arabic or English speaking, must proceed through a log-in procedure. The latter adheres to the security facilities implemented in the operating system and consists of entering a log-in name ("user name") and one or more passwords. Security procedures also exist in the operating system for the system administrator to preclude access by specific categories of users to certain system facilities; for example, data entry personnel can be restricted to appropriate utilities, external users can be allowed access to a subset of databases only, etc.

Following successful log-in, the user is invited to select, from a menu screen, one particular function supported by the STI Network. The principal functions that the STI Network proposes to support at this time are described below; they are: database building and management, database search and retrieval, file creation and processing, communications, and system administration.

1. Database Building and Management

This function is restricted, at least initially, to the staff of the information service node. The function encompasses the generation of new databases (both those compiled by the node or created by reformatting leased databases), updating and maintenance of databases, and database withdrawal or purging. Typically, database maintenance functions will be carried out by the system administrator at times other than peak system usage.

In principle, the STI Network nodes are not inhibited from permitting selling computer facilities to external users (e.g., the host organization or other clients) and allowing them to build their own, "private" databases. When making facilities available for private database building, however, the ISN must consider that this will tie up disc space, one or

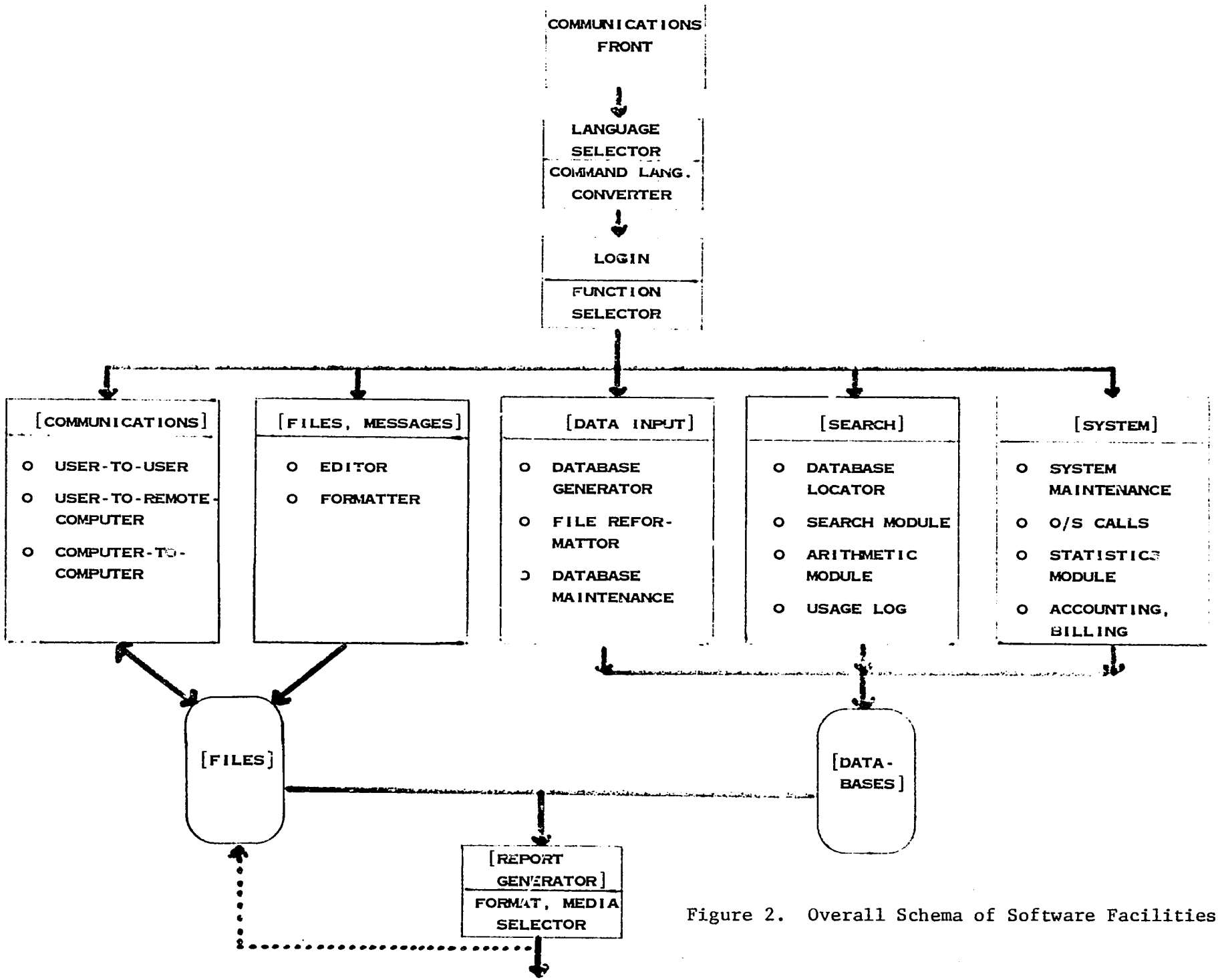

Figure 2. Overall Schema of Software Facilities.

more ports as well as consume CPU time. It is not recommended, therefore, that this service be made available in the first year. In any case, it is absolutely essential that selling computer time and facilities does not impair, in any manner, the public information services that are the primary function of the STI Network.

2. Database Search and Retrieval

This is the primary user-oriented function of the STI Network. Users accessing particular databases may be external users, either casual or staff members of other nodes; and professional database searchers of the local ISN. The latter will account for the bulk of early system use, but major user organizations (e.g., university or large corporate libraries, research institutions, and the like) should be encouraged to seek remote access through dial-up or leased telephone lines as soon as the ISNs have databases of interest to them.

3. File Creation and Processing

This capability allows the user to create and manipulate bodies of electronic data called files. A file may be a letter, a document, a message for electronic mail, the stored output of database search, a computer program, or any string of alphanumeric data.

The powerful file processing capabilities embodied in the operating system of the STI Network computers include facilities for document editing, formatting, and printing or typesetting. STI Network nodes would be well advised not to make these facilities generally available to the general public, or else they may find that the remote use of their computers is for text processing rather than for database search and retrieval.

The principal users of the file processing software are the node staff. Apart from correspondence applications (such as billing, mass mailing, memoranda, newsletters), these facilities will be used for the generation of printed products from databases: directories, various tabulations, technical reports prepared for customers, and the like. STI Network plans call for the publication of a national abstracting journal of Egyptian scientific and technical literature; this too will be generated via the file processing capabilities of the operating system.

4. Communications

This capability comprises three services: user-to-user (electronic mail), user-to-computer, and computer-to-computer communications.

a) Electronic Mail. Electronic mail is a message-forwarding service from one user to one or more other users [94]. This capability allows any logged-in user with permission to use the "mail" or "write" commands to:

send ("mail") messages or files to the mailboxes of one or more specified users on the same or any other computer in the Network; to read mail received to his own mailbox from other users; and to "write" (interactively correspond) with another user who is logged in. This capability also allows the system administrator to write (send messages) to the terminals of all users currently logged in on a particular computer. Electronic mail is embedded in the operating system of each of the Network machines.

b) Computer-to-Computer Communication. This capability provides means of communication between and among the computers of the National STI Network. Using (initially) dialup telephone lines, the system administrator or any user having the appropriate permission may call up one or more computer systems on the STI Network and carry out a range of functions. The latter include transfer of files and operations on remote files (such as merging or updating remote files, program execution, and similar operations). Strict control must be maintained by the system administrator at all times regarding the permissions to access a node's own computer resources.

c) User-to-Remote-Computer Communication. Apropriate automatic facilities exist at each STI Network computer for ISN professional personnel to access, via available telecommunications lines, commercial databases outside the STI Network, both in Egypt and abroad [63]. Accessing foreign databases is a question of the availability, quality and economics of telecommunications resources. Until Egypt implements access to international packet-switching networks, online searching of foreign databases must utilize long-distance telephone lines; this invariably doubles or triples the average search cost, and it is likely to be used infrequently.

Each information service node computer initially supports three communications ports at 300/1200-baud rates. This facility provides an integral, full-duplex, dial line modem with automatic originate/answer transmission and reception over switched or leased telephone networks, in either 300 or 1200 bits per second; in the answer mode the modem automatically recognizes and accepts either incoming data rate.

High speed computer telecommunications for bulk data transfer are not supported by the STI Network at this time. While each computer system has the appropriate software facilities for such communication, the Egyptian telecommunications network is not suitable for high speed data transfer at this time. Similarly, computer-to-computer communications between a machine of the STI Network and a non-Network computer is not supported automatically, although facilities exist in the UNIX operating system to establish a telephone connection with suitable non-Network machines.

5. System Administration

The system administration utilities are provided partly through the operating system, partly through the database management software. The operating system administration utilities allow the computer system

administrator of each ISN site to perform the following functions: management of the file system and other utilities (file system consistency and size checks, debugging and repair); machine and media management (computer power-up and power-down, file installation, backup, removal, etc.); and user services management (adding and deleting users and user groups, establishing passwords and access permissions, and administering computer security features).

The system administrator's responsibilities include control over software modifications. As a matter of principle, any modifications of the operating system and of the standard database management system used by the STI Network should be permitted and installed only on a Network-wide basis. Local enhancement of software should be allowed only if it does not corrupt the standard operating and DBMS software; if possible, it should be limited to shell scripts; and it must be fully documented. Any attempts to modify unilaterally Network-wide standards and compatible resources (e.g., record format, query language, output formats) should be strongly resisted. On the other hand, local enhancements that do not corrupt these standards but are built upon them are welcome.

The database management system software administration capabilities further provide facilities for administration of public use of the system, including monitoring of user activities, and service billing and accounting.

Access to system administrative functions is restricted to individuals specified by the ISN manager.

D. EGYPTIAN STI NETWORK DATABASES

This chapter describes the organizational approach recommended to develop a pool of public-access databases in Egypt; discusses the content of these databases; and makes observations regarding the nature of this content. The latter point has an important bearing on the database management software to be provided.

1. Development of Sectoral Database Pools

Figure 3 shows four distinct sources contributing to the development of databases in the STI Network. They are:

a) Sectoral ISNs. Each sectoral node must anticipate assuming responsibility for the establishment of public databases relevant to problem solvers in its sector. The ISN may generate such databases alone as an inhouse effort, or jointly with other sectoral agencies or organizations. The ISN will remain responsible for the continuous maintenance and updating of databases generated under its auspices.

b) Other Egyptian Agencies. Any ISN may consider mounting on its computer databases whose contents come from data banks of other Egyptian organizations, including other ISNs. Candidates for sources of such databases are public agencies that collect extensive data as part of their mission, and are prepared to release regularly portions of the data that is useful to the public. A resourceful ISN should be on the lookout for repositories of quality information in agencies of its sector, and negotiate agreements to have such information (or subsets thereof) mounted on its computer as an ISN database.

c) United Nations. The UN and its specialized agencies are a source of prolific data, of national and global coverage. Whereas much of this data is not available in machine-readable form, and some is restricted, valuable files are available from these organizations at royalty-free cost [96]. Table 2 lists names of UN organizations that maintain databases of various sorts.

d) Foreign Database Vendors. Commercial databases available on the international market number well over 1,000, and many may be subscribed to in magnetic tape form, for a fee. An STI Network node considering the subscription to a foreign database should be guided principally by economic considerations: as a rule, subscription should be investigated when the demonstrable demand for using a foreign database is such that the cost of searching it overseas approaches the cost of database subscription.

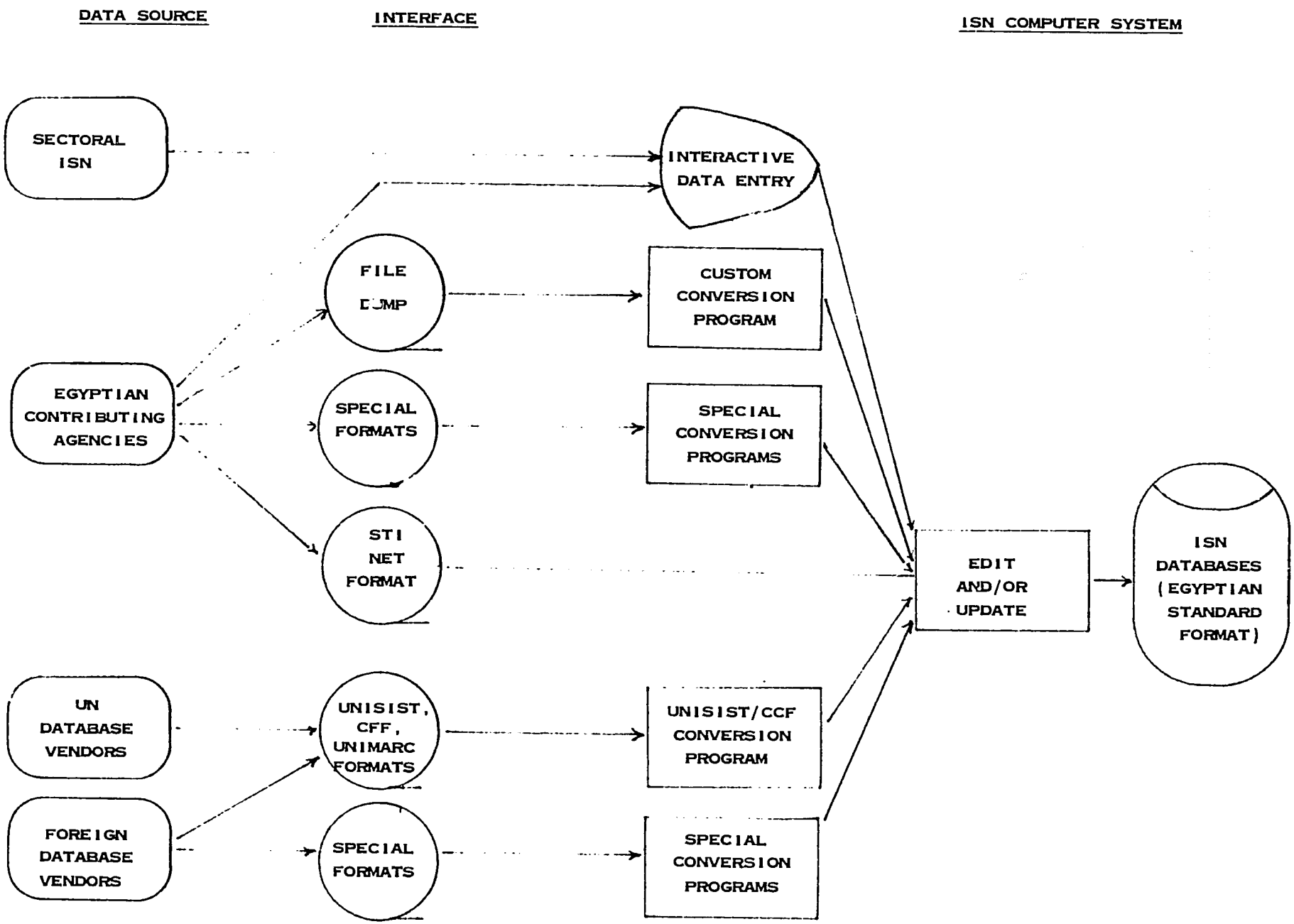

Figure 3. ISN Databases: Input.

Table 2. United Nations Organizations Operating Information Systems.

UN	United Nations
ECE	Economic Commission for Asia and the Pacific
ESCAP	Economic Commission for Latin America
CEPAL	Economic Commission for Latin America
ECA	Economic Commission for Africa
ECWA	Economic Commission for Western Asia
UNDRO	Office of the United Nations Disaster Relief Coordinator
WFC	World Food Council (UN/FAO)
UNCTAD	United Nations Conference on Trade and Development
UNIDO	United Nations Industrial Development Organization
UNEP	United Nations Environment Programme
UNICEF	United Nations Children's Fund
UNDP	United Nations Development Programme
WFP	World Food Programme (UN/FAO)
UNHCR	Office of the United Nations High Commissioner for Refugees
UNRWA	United Nations Relief and Works Agency for Palestine Refugees in the Near East
UNITAR	United Nations Institute for Training and Research
UNU	United Nations University
ITC	International Trade Centre (UNCTAD/GATT)
ICJ	International Court of Justice
ILO	International Labour Organisation
FAO	Food and Agriculture Organization of the United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
WHO	World Health Organization
WBG	World Bank Group
IMF	International Monetary Fund
ICAO	International Civil Aviation Organization
UPU	Universal Postal Union
ITU	International Telecommunication Union
WMO	World Meteorological Organization
IMCO	Inter-Governmental Maritime Consultative Organization
WIPO	World Intellectual Property Organization
IAEA	International Atomic Energy Agency
GATT	General Agreement on Tariffs and Trade
IFAD	International Fund for Agricultural Development

Figure 3 shows, furthermore, that with the exception of the data collected by the sectoral ISN (which can be entered into sectoral databases directly and interactively via a terminal), data provided to the nodes in machine-readable form (normally as a flat ASCII file on industry standard 9-inch magnetic tape) is likely to come in one of several format types: a file dump; a standard record format, where "standard" refers to the Egyptian STI Network record format or one of the following internationally agreed upon formats: UNISIST [18], UNIMARC [39], or the Command Communications Format [82]; or a variety of special formats, each being specific to a particular database (such as the Engineering Index). Each of these format types will require conversion to the standard Egyptian STI Network record format, an optional editing step, and the database update step.

For purposes of completeness, Figure 4 shows the output interfaces of the ISN computer system, including several popular formats for international exchange of machine-readable files.

Understandably, the database software selected for the Egyptian National STI Network must have facilities for an easy conversion to and from the standard database format of the Network.

2. Contents of Sectoral Databases

As a public information utility, the Egyptian STI Network is principally geared to assist the Egyptian public with provision of data and information required in problem solving activites. Existing theoretical models of the interaction process between problem solving and information are too complex for the needs of this document; we have therefore developed a simpler, generic model (Figure 5) that depicts the general stages of the process. The model is based on work of Polya, Newell and others.

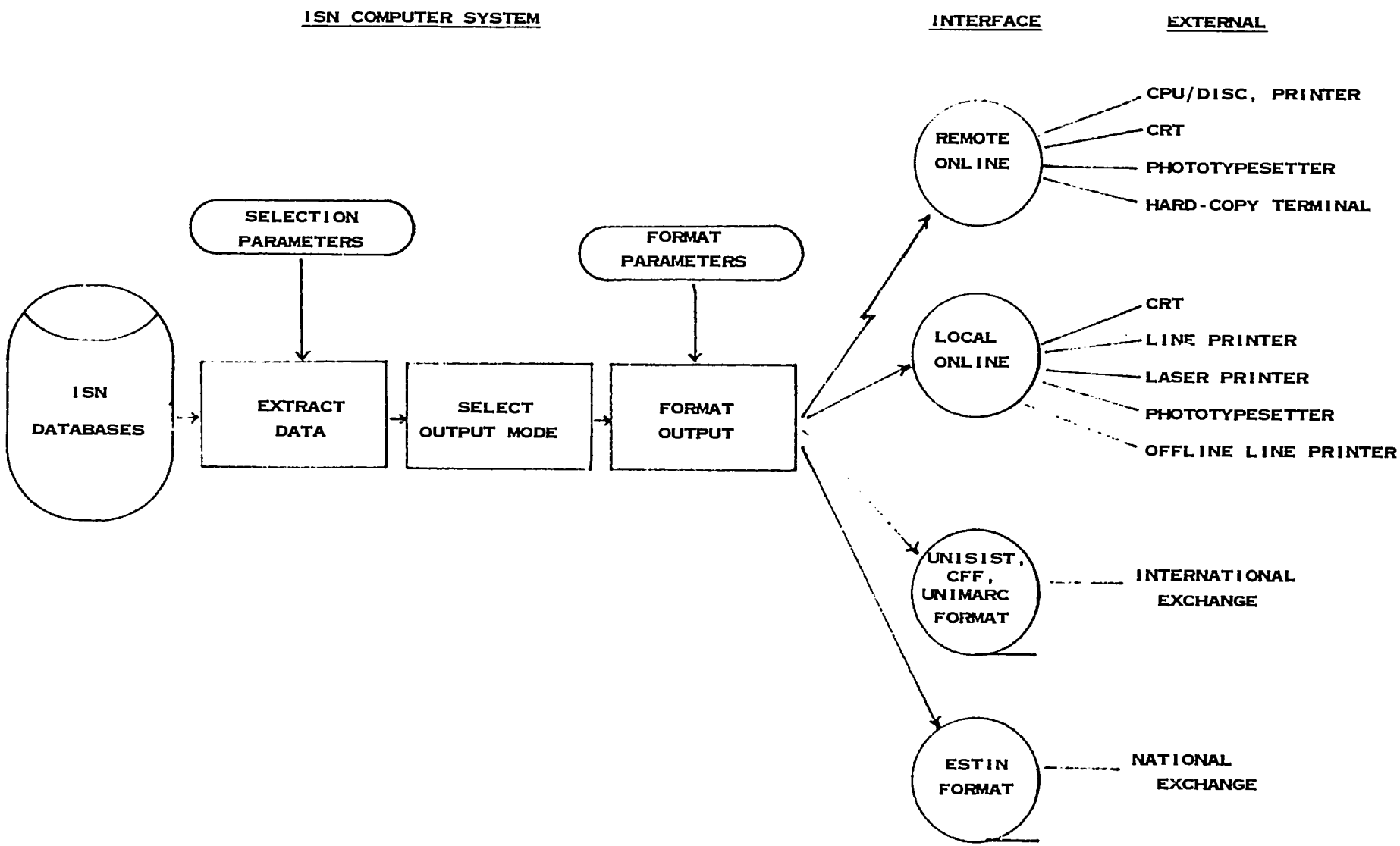

Figure 4. ISN Databases: Output.

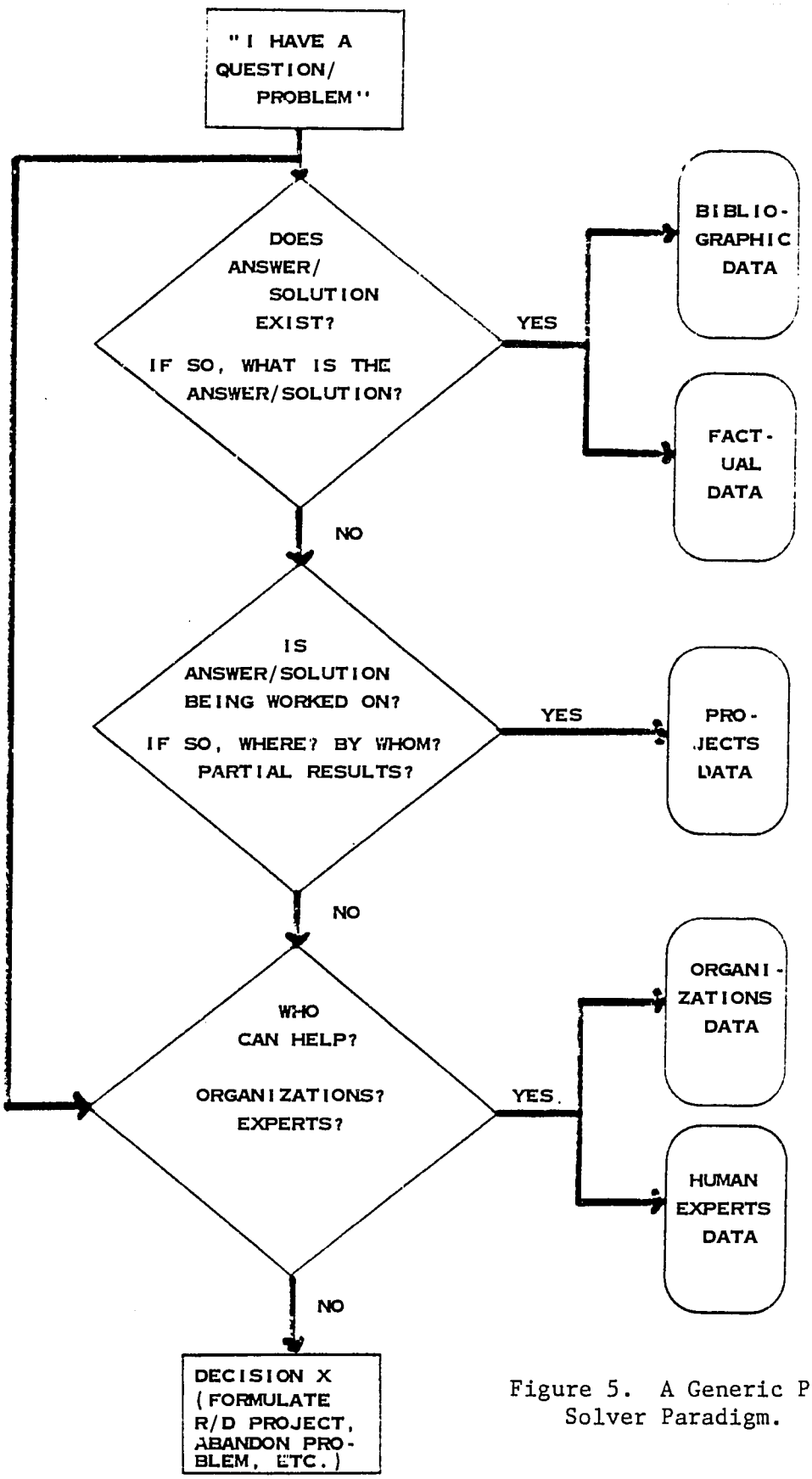

Figure 5. A Generic Problem Solver Paradigm.

At each stage, effective aid to problem-solving can be provided through data/information support. The major categories of data/information resources that cumulatively meet the information-support function in problem-solving are: bibliographic, managerial, organizational, and personal. "Managerial" resources are typically exemplified by data on formal activity plans, such as research projects.

The design phase of the STI project recommended that the STI Network initially undertake, in each sector, the organization of four data/information resource categories: bibliographic, managerial, organizational, and personal. The managerial resource data is to be initially restricted to R&D projects. The outcome of this effort will be a computer-based utility consisting of public data about Egyptian organizations and agencies, their activity plans, the expertise of individual persons, and the record of Egyptian scientific and technical writings in five sectors: agriculture, energy, medicine and health care, industry, and science/academia. The rationale for this recommendation included the following reasons: these four data categories are relevant to all five sectors; the data accumulated by the respective sectors can be aggregated into multisectoral, quasi-national databases; and a single record structure probably can be developed to accommodate all four types of data and information, thus greatly simplifying both the creation and the use of these databases. The four database types are typically the first to be built by national information systems.

The multisectoral bibliographic database is a particularly important product; it combines surrogates of different types of Egyptian scientific and technical (S&T) literature. The types of literature initially included are Egyptian S&T books, periodicals, conference proceedings, technical reports, theses, patents, and standards. The generation of a national S&T bibliographic database will be a carefully orchestrated effort of a number of agencies, including some or all sectoral information service nodes. Using common record format these agencies will provide regular flow of bibliographic item descriptions; these will be converted to machine-readable form at selected sectoral information nodes, and eventually merged at one ISN into a single national bibliographic database. Copies of the database can be made available to interested sectoral nodes; and based on it, a national abstracting journal of Egyptian scientific and technical literature will be published.

A fifth category of data-gathering was also identified as being of urgent interest -- factual data such as inventories of manufactured goods, farm products, energy prices, etc. This data category is sector-specific, and normally not intended for multisectoral aggregation [62,73]. Each factual database is likely to have a different record structure. Thus while the sectoral information system software must be hospitable to building factual databases, their development must be reviewed and determined individually by each sectoral node. By the way of an example, Table 3 lists factual data of interest to the energy sector.

Table 3. Factual Data: The Energy Sector*

Energy Consumption

Sectoral Energy Use

- o annual
- o sector and subsector
- o fuel use and useful energy

Energy Consumption Equipment

- o equipment by sector (industry, transportation, agriculture, etc.)
  - fuel use
  - fuel switching capability
  - equipment (size, age)

Energy Resources

Resources by type, status (known, proven)

International Resource Availability

- o type
- o cost
- o transportation

Energy Production

- o fuel and resource production
- o annual

---

*Courtesy Office of Energy Planning (Cairo)

3. Nature of Data in Databases

A closer look at the four text-oriented types of databases discussed above reveals that they all contain data elements comprised of sentences and natural language aggregates; and at least one of them (the bibliographic database) will contain unformatted, free text. In contrast, factual databases are likely to contain relatively structured data, of both numeric and alphabetic types.

One major conclusion drawn from this observation is that the database software of the Egyptian STI system must possess attributes traditionally associated with information retrieval systems -- i.e., the ability to operate on bodies of full-text natural language. Database management systems that handle effectively structured data types but do not possess facilities for handling unformatted text are not adequate for the purposes of the STI Network.

E. STI NETWORK DATABASE SOFTWARE REQUIREMENTS

This chapter discusses the database software requirements and characteristics of the Egyptian STI Network. The literature provides a variety of guides which discuss desirable characteristics of database software [4,16,20,36,43,48,62,71,74,77,84,89,101].

For our purposes it is propitious to divide the database software requirements into two broad groups. The first, operational requirements, refers to environmental and performance factors; the second, functional requirements, refers to capabilities that the database executes. The table below summarizes the main sets of factors falling under these two groups.

Operational	Functional
National-standard	Database definition
User orientation	Input/maintenance
Performance	Search/processing
Security	Output
Modifiability	Administration
Support	

Whenever possible in the following discussion, an attempt has been made to distinguish between mandatory characteristics and those that are either desirable or optional.

(OPERATIONAL REQUIREMENTS)

1. National-Standard Features

The database software of the Egyptian National STI Network should be capable of being viewed and adopted as a de facto standard. The intention of this statement is not to propose that the national standards agency declare the software to be a national standard; rather, the intent is to have this software commonly adopted and used by the STI Network and, possibly, other public-sector information systems.

The principal features required to make the software widely acceptable are the following:

a) Portability. Portability can be attained in more than one way; in the circumstances of the STI Network it means that the database software must run, as a minimum, under the UNIX operating system. The latter was selected as the standard O/S of the STI Network, partly because of its own portability.

b) Bilingual Mode. The software must support the building of databases in English and Arabic, and it must accommodate Arabic-speaking database users. It is recognized that the Arabic language capability will have to be developed under a separate effort, inasmuch as off-the-shelf bilingual database software is rare. Accordingly, the software copyright owner must be amenable to a joint effort to develop the Arabic capabilities.

c) Common Command Language. To the greatest extent possible, the query language should employ common commands for Network-wide interaction with the databases. The set of commands nearest to a standard is the Common Command Language (CCL) of EURONET [61]. CCL is intended, however, for bibliographic database retrieval, and may not be compatible with DBMS query languages. As a minimum, the database software of the STI Network must allow the implementation of a CCL-like set of commands; facilities for command renaming ("alias") are therefore essential (they are also useful for creating Arabic commands).

d) Common Communications Format. The record format of databases of the STI Network remains to be developed. Whatever the format, however, uncomplicated software facilities must be available to reformat the databases into "standard" communication formats, for purposes of exchange. Bibliographic record formats that enjoy internationally a near-standard status are UNIMARC [39], the UNISIST record format [18], and the very recently released Common Communications Format [82]; other, more restricted formats also exist [1,38]. (As regards other UNIX applications software, the common interface newly adopted is the Independent Software Information Standard developed by the ISIS Group, a consortium of leading UNIX software vendors in the U. S. The ISIS data interchange uses flat, unformatted ASCII files with a standard information header and data format.)

2. End-User Orientation

Database technology has advanced to the point at which it serves directly -- without a programmer intermediary -- the occasional ("casual") user who is not expert in computer use. This ease of use by casual users is a mandatory characteristic of the software to be selected for the Egyptian STI Network. The software features that characterize this property include the following:

a) Unattended Operation. It is mandatory that the software be a mature product running unattended except for assistance to novice users, and for functions performed by database managers.

b) Friendly Interface Language. It has been amply established [7,22,42,57,80,81,87] that binary choice (yes/no), multiple choice (menu), and "fill-in-the-blank" techniques of interaction are easy to learn by occasional database users, and are much preferable to low-level programming languages which require the user to specify how a search is to be carried out. If the selected software does not have these techniques built-in, it must provide easy-to-use utilities to generate screens, forms, output formats and the like.

c) "Help" Assistance. During interactive sessions with a database, users are expected to make various choices. Since the criteria of choice, or the menu of selection of items, may not be obvious to the user, the "help" command provides an online clarification or explanation. "Help" commands have other functions as well -- e.g., succinct description of user commands, in the form of an online manual. A mature, end-user oriented database software system contains such "help" features as a mandatory characteristic [35].

3. Performance

This class of characteristics addresses the ability of software to operate robustly and reliably in an environment of users demanding high-quality, secure and rapid services. Many of these characteristics are related to program architecture and other techniques used by the software (e.g., space management); rather than being concerned with these techniques per se, our interest is in the overt behavior (performance) of the database system software. The principal characteristics that the selected database software must possess include the following:

a) Expandable Multi-User Operation. The software must support concurrent database operations by multiple users working with one or more databases simultaneously. As user demand increases and appropriate physical facilities are added (CPU memory size, disc channel capacity, etc.) the software must continue to accommodate increasing numbers of simultaneous users. A single-user database software is unacceptable; the existence of a fully compatible single-user version of the multiuser DBMS for personal computers is desirable, however.

b) Database Size. The software must support the following database size parameters:

Entity	Mandatory	Desirable
No. of databases (relations)	20	no limit
No. of records in database (relation)	0.5 million	2 million
Size of record (relation)	2,000 bytes	65,000 bytes
No. of record fields (attributes)	50	250
Size of field (attribute)	2,000 bytes	10,000 bytes

The mandatory data volumes exceed substantially the resources of the initially supplied computer system. For example, a single database of 0.5 million records, each record averaging 1,000 bytes, and its full-text indexes will require disc storage capacity of 1,000 Mb, or roughly three times larger than the disc supplied initially. The robustness of these sizes guarantees that the software will be usable when the storage technology provides for much larger capacities at considerably lower cost, as is expected with the optical disc medium. The Egyptian STI Network will then be able to switch to these technologies and continue using the same database software.

It should be noted that Egyptian sectoral databases will be much smaller than the example above.

c) Response Times. Clearly, in a multi-user environment, response time is a function of complex factors including hardware characteristics, disc access speed, and the like. User response time is also considerably affected by the file structure, the indexing techniques, and by file searching algorithms used by the database software. Finally, response time depends on the nature of the database content, and whether the search is conducted for specific items of known or unknown key values. The combinations of these parameters are so complex and uncertain that it would not be propitious to prescribe requirements in terms of mandatory file structure, file searching algorithms, indexing techniques, and other parameters.

The design study produced a guideline for desirable response times, based on the existing state of the art and on what professional users of information systems consider to be fully acceptable response. These guidelines should perhaps be viewed as desirable, rather than mandatory, when candidate software is evaluated empirically. The guidelines recommended the following response times:

Type of Online User	Response Times (seconds)
Inquiry/Response*	2-20
Data Entry**	2
Administrative***	5

*database searching, script preparation (depending on process complexity)
**database development (online)
***maintenance of administrative files, billing, training

d) Data Compression. With large databases, data compression techniques are common, as they reduce the needed capacity of expensive storage media. For full-text databases, compression of the textual file is mandatory; of ancillary index files, desirable. The compression algorithm efficiency should be such as to compress the file by 40% or more. The database software is not expected to possess a compression capability for Arabic data; the latter may have to be developed by Egyptian computer scientists.

e) Recovery Procedures. A system failure can be caused by hardware problems, certain software problems and by power outages. Recovery procedures implemented in the database software are necessary to make certain that after such a failure the database is able to restart quickly and that the data is consistent. Data errors may be logical (e.g., updates that were not made, or were made incorrectly); or physical (caused by hardware and media malfunctions). It is mandatory that the database software contain effective procedures for recovery from logical and physical failures. The preferable approach to recovery is by recording all modifications made to data in a recovery log; however, other effective approaches are also acceptable.

4. Security

a) Access Control. The database software has to have mandatory provisions for controlling various database functions and their accessibility to different categories of users, so as to shield facilities and data that are irrelevant to a specific user. The definition of such "user views" is to be performed by the database manager.

Secondly, mandatory facilities must be available to grant or deny access permissions by users to files at the level of the database (relation), record (tuple), and field (attribute).

It is desirable that the security techniques of the database software utilize the login and password utilities of the UNIX operating system.

b) Concurrency Control. It is desirable that qualified users be able to concurrently query and change (insert, update, delete) a database. It is mandatory, however, that the system allow only a single user to change data in the database; a locking facility must be available to minimize interference between concurrent users and to prevent access to inconsistent data.

5. Modifiability

Some of the specifications of the Egyptian STI Network services -- e.g., the bilingual user interface -- very likely exceed the facilities of even the most sophisticated database software on the market. The STI Network must therefore anticipate the selected software to require nontrivial extension, to be carried out as much as possible in Egypt. This effort can be accomplished reasonably efficiently if Egypt is provided with suitable software tools [47]. The latter must be an integral part of the database software; in theory it is possible to employ an external set of tools (such as a third-party applications generator software) but only with difficulty.

To suit in the Egyptian environment, the selected database software therefore should have the following characteristics:

a) Programming Language. It is desirable that the software be written in C, the native programming language of the operating system UNIX. If not written in C, it is mandatory that the software be highly portable, and that it run smoothly under the C compiler.

b) Host Language Interface. The UNIX command interpreter provides powerful yet simple-to-use facilities for writing concise computer programs or subroutines called "shell scripts." While shell programming is not a skill expected of a non-computer specialist or casual user, it becomes a highly productive tool in the hands of a relatively junior programmer who wishes to implement customized operations on the database. The database software must provide an interface to the facilities of the UNIX operating system, and to the C language of UNIX.

c) Customizing Utilities. It is mandatory that the database software have facilities to easily create and modify screens (CRT display frames) and forms (output displays and printouts). A language for this capability must be provided as an integral part of the software [47]. The users of this facility are to be the database managers or programmers, not casual end-users; therefore, this facility needs to be provided only for English-speaking computer professionals. This notwithstanding, the customizing utilities will be employed to generate Arabic-language screens and forms as well.

6. Support

This category of environmental specifications seeks a guarantee that the database software is a mature product, and that its vendor is a mature enterprise prepared to provide Egypt with the commitment and support relevant to operating the STI Network. The following requirements apply:

a) Support Responsibility. The Egyptian STI Network nodes initially expect to have a single maintenance contract with the hardware vendor (AT&T) under which this vendor will be responsible for the maintenance of hardware, the operating system, and the database software. Since the database software will not be an AT&T product (AT&T does not, at this time, offer a DBMS software of its own make), an agreement must exist between AT&T and the prospective datatabase software owner giving AT&T the necessary rights to lease and service the software package. Furthermore, an agreement must exist between AT&T and the database software owner giving AT&T access to the latter's information and expertise, so as to be able to perform the maintenance function properly. In the absence of such agreement, the STI Network (or the Georgia Institute of Technology) must negotiate an appropriate contractual relationship with the database software owner, making the latter responsible for maintenance of the software.

b) Software Owner Support. Regardless of who contracts with the database software owner, it is mandatory that the latter commit himself to assisting the Egyptian STI Network with the Arabization and other extensions of the database software. Such a commitment presupposes arrangements regarding the software copyright ownership. While it is not expected that the software owner will release the copyright of the bilingual software, it is mandatory that the licensing arrangements account for the contribution by Egypt or its consultants to the software extensions, unless such a contribution was financed in full by the software owner.

The software owner must make available to Egypt (or the software maintenance contractor) expert staff knowledgeable in the software internals and code.

c) Training. It is mandatory that the database software owner have available instructors and training materials to train, on-site in Egypt, STI Network technical personnel in the use, administration and maintenance of the software, unless the same responsibility is assumed by AT&T. The implementation of such training as a self-instruction package under the UNIX Instructor Workbench is desirable. The database software owner has to be amenable to have Egyptian instructors assume the training function as soon as possible, and he has to agree to provide them with training materials, manuals, CAI packages, and other aids as these become available or are updated, at nominal costs.

d) Documentation. Another mandatory element of the database software support is full documentation, consisting of high-quality user and administrator manuals. The manuals should be made available at nominal

prices, or else be allowed to be reproduced at cost by the STI Network. It is mandatory that permission will be granted to translate or rewrite the manuals into Arabic and publish them in Egypt. Program documentation (source code) must be available to those responsible for software maintenance.

e) Local Assistance. Large database vendors in the industrialized countries provide telephone assistance to users encountering difficulties in using their facilities. It is desirable that the software owner assist the STI Network in setting up and training Egyptian "online user assistance" personnel, and to share the software owner's expertise with clientele problems.

(FUNCTIONAL REQUIREMENTS)

7. Database Definition

a) Database Creation. It is mandatory that easy-to-use facilities be built into the software allowing the database manager to define and structure a database and its logical record. The facilities should be of the fill-in-the-blank or menu form, and should guide the database manager to i) assign the names of the database, primary key(s), and record fields or attributes, ii) specify data types, iii) designate each field or attribute as fixed or variable, and iv) stipulate the length of the field.

b) Data Types. The following data types must be supported on a mandatory basis by the software: i) unformatted, variable-length text data; ii) structured text data (e.g., personal names); iii) integers (numeric data of reasonable size), iv) date; v) dollar; and vi) floating point.

c) Indexable Fields or Attributes. The database manager must have the ability to decide, for each field of the record, whether or not it is to be indexed.

d) Access Permission. The database manager must have the ability to assign access control codes at the levels of record field, record, and database.

8. Database Input and Maintenance

The software must have the following mandatory functions:

a) Data Entry. Data entry via terminal keyboard should be assisted with prompting (typically, the field name), and be of the fill-in-the-blank type. The following functions must be supported, as a minimum: i) automatic verification of data type correctness, ii) automatic check of allowable data length, iii) automatic check of the forms of date and

ccession number; and iv) display of suitable error messages if the conditions are not met. Simple means (command) should allow the automatic insertion of repeated fields in successive records. Other desirable features include verification of certain field formats during data entry via "masks".

b) Data File Loading. The software must support the loading of data files prepared offline, whether or not such datafiles are in the record format of the database. It is mandatory that facilities exist to load flat, uncompressed ASCII files directly into the database format.

c) Data Indexing. The database software must have a mandatory capability for automatic machine indexing of full text (free-text indexing). Manual controlled-vocabulary indexing must be possible (that is, human indexers entering into data fields indexing terms chosen from a controlled vocabulary or thesaurus). Maintenance of the controlled vocabulary or thesaurus as a separate database is a desirable feature.

c) Database Changes. Capability must exist for changing (modifying, deleting) database content at the record (tuple) and field (attribute) levels.

9. Database Search and Processing

This capability refers to processing users' search statements against the database. The mandatory and desirable functions are as follows:

a) Query Language. The software must employ a high-level language for interactive database search, such that the user need not be concerned with the issue of how the search is carried out. Mnemonic commands with arguments, fill-in-the-blank, multiple choice (menu), and QBE (Query-By-Example) [105] are all acceptable techniques for text searching and retrieval, and preferable to the SQL (Structured Query Language) syntax. The latter is preferable, however, for retrieval from factual databases.

b) Search Fields. It is mandatory that the user have the option of selecting the fields (attributes) on which the search should be performed (in SQL, such choice is implied).

c) Search Capabilities. The mandatory capabilities for database query processing include: i) searching on strings, adjacent strings, strings with the same sentence, word roots, and truncated words (both left and right truncation); ii) boolean operators AND, OR, and NOT; iii) range operators =, >, <, =>, =<; and key statistical operators (sum, average, count, percent, distribution).

d) Search Aids. Mandatory online search aids include: i) for free-text searching, display of strings alphabetically preceding and following the keyword, with frequency of postings; ii) for controlled-text searching, display of appropriate entries from a thesaurus or an authority list (if

maintained in machine-readable form); and iii) display of postings for each search term. More sophisticated aids are desirable, particularly aids that assist users in recalling synonyms and related terms and thereby refine search strategy (such as the "zoom" capability of the European Space Agency's information retrieval software ESA-SEARCH).

e) Search Query Storing. Users must be able to file away the query for later use (sometimes called SDI search).

f) Multiple Database Search. For relational software, this is a mandatory capability; for software based on indexed file structure, a desirable capability.

10. Search Output and Processing

The following features are considered mandatory:

a) Output Redirection. The user shall have the options of i) viewing the output on the terminal screen, ii) spooling it to the central facility's line printer for offline printing, iii) downloading the output to another computer or a remote terminal/printer; and iv) saving the output in the form of a file on the central computer of the node.

b) Output Content Selection. A menu of standard output formats available to the user is mandatory; included in it must be standard tabular and bibliographic formats. Also mandatory is the ability for the user to select record elements that will comprise the output. Custom-tailoring of output format is a desirable feature.

11. Administration

Facilities must be available to perform administrative functions related to database use, as follows:

a) User-Available Facility. Each legitimate user must be able to request, online, statistical information relating to his last interaction (length of time, resources used, resources spent).

b) Administrative Facilities. The software has to have mandatory programs to calculate and generate system usage statistics, user accounting and billing. Preferably, administrative utilities available in the UNIX operating system should be relied upon.

F. SURVEY OF COMMERCIAL IR/DBMS SOFTWARE

This chapter reviews the steps taken to survey and analyze commercial software systems for potential use in the Egyptian National STI Network.

Once the mandatory and desirable specifications of the information system software for Egypt have been identified, the steps involved in the selection of a system are relatively straightforward, and several guides exist to aid the selection process [36,43,48,65,71,73,79,85,102,104]. This study has employed the following sequence of actions: a) survey and collection of information on existing IR and DBMS software systems; b) study of existing, previous comparative evaluations; c) preliminary review of existing software packages aimed at eliminating software that does not meet mandatory specifications; d) in depth study of candidate software; and e) preparation and documentation of recommendations. These steps are briefly explained below.

1. Database Software Survey

A number of sources were studied in the process of obtaining a list of commercially existing databases. These sources include:

- Searching databases of commercial vendors (Dialog, BRS), paticularly COMPENDEX, NTIS, SOFT, INSPEC, and the Computer Database;

- Commercial software directories (Auerbach Computer Technology Reports, Datapro Directory of Software, Computer Program Abstracts (NASA/COSMIC), The Federal Software Exchange Catalog (NTIS), Catalog of Directories of Computer Software Applications (NTIS), International Directory of Software (CUYE Publications), Data Decisions Software, Data Sources, ICP Software Directory, Elsevier Software Catalog; UNIX Software Guidebook (DEC), and Onager's UNIX Software Directory);

- Private and electronic bulletin boards and user network clubs (Edunet, CSNet, The Source, The Compuserve Information Sources, ICST Microcomputer Electronic Information Exchange, ICST Computer Performance Evaluation Users Group, the Federal ADP Users Group, and the UNIX Users Group);

- Appropriate conference proceedings (National Computer Conference, ACM Annual Conference, ASIS Annual Meeting, annual Online Meetings);

f) Professional news literature (ComputerWorld, Mini/Micro, Datamation, Electronic News, Online Journal);

- Personal contacts in and outside the United States.

Inquiries have been directed at identified software vendors and product literature, manuals, demonstrations, etc. requested or procured. This survey has been conducted over a period of 24 months, inasmuch as new database software keeps arriving on the market.

Information has been assembled on nearly 100 commercial database products. After elimination of obviously irrelevant products (such as simple report writers, transaction processors, etc.), the final list of products studied is shown in Table 4.

2. Previous Comparative Studies

Several commercial and private comparisons and evaluations of database software were identified, acquired and studied [2,14,16,26,32,44,47,49,52,62,64,67,79,86].

3. Preliminary Review

When matching the mandatory specifications of the Egyptian STI Network's database software against the main features of the packages listed in Table 4, a substantial number of the latter were quickly eliminated as unsuitable. Typical reasons for excluding these packages from further consideration were: a) restriction to single-user, personal computer systems, b) inability to obtain technical information from the vendor, c) implementation under operating systems other than UNIX, and d) lack of features qualifying the systems as mature, full-capability data and/or information retrieval software.

4. Analysis of Candidate Software

This critical and time-consuming step was based on the following strategy. Given the database software requirements of the Egyptian STI Network, we first attempted to identify and study software packages capable of storing and retrieving both text and data types. Not having found a suitable candidate, we proceeded to study selected relational DBMS software, with the objective of identifying packages capable of, or easily amendable to, handling information retrieval applications. Failing in this goal, we proceeded to evaluate available information retrieval software.

The analysis of each software package consisted in comparing its features (as available from documentation such as manuals, and telephone discussions with vendor or client) against the mandatory specifications described earlier.

The findings and recommendations are contained in the next two chapters. The supporting documentary materials assembled in this evaluation study are being made available to Egypt under separate cover.

Table 4. List of Reviewed Software

SOURCE	DATABASE	VENDOR
2	10-BASE	Fox Research Inc.
6	ADABAS-M	Software AG
4	AMBASE	Amcor Computer Corp
5,6	BASIS	Battelle Columbus Labs.
6	BIBPRO IV	Informatics Inc.
1,3,5,6	BRS/SEARCH	BRS Inc.
4	BUSINESS DB	Charles Mann & Assoc.
4	CAIRS	Leatherhead Food RA
4	CONDOR SERIES 2	Condor Computer Corp
4	CREATE	Complete Computer Systems
4	DATA 80	The Software Store
4	DATA BOSS 32	Florida Computer, Inc
3	DATA-BASE PLUS	Tominy, Inc.
4	DATABASE 2	IBM Sta Teresa Lab
2	DATAEASE	Software Solutions, Inc.
4	DATAHANDLER	Miller Microcomputer Serv
4	DATASCAN	Data Management Systems
2,4	DBASE II	Ashton-Tate
6	DBMS	Prime Computer Inc
1	DBMS III	ISE-USA
6	DBMS-11	Digital Equipment Corp
6	DBMS-990	texas instruments inc
4	DBQUERY	Gemini Information System
6	DG-DBMS	Data General Corp
4	DRS	Advanced Data Management
6	ENCOMPASS	Tandem Computers Inc
1,5	ETIS	US Army, Constr'n Eng
4	FACTMATCHER	Mini-Computer Systems Inc
1,5	FINDER	Aaron/Smith Assoc. Inc
6	HARRIS AZ7	Harris Computer Systems
1,5	IDM-500	Britton-Lee Inc.
5,6	IDOL	Science Management Corp
6	IMAGE	Hewlett-Packard Co
1	IMPRS	Ruf Coorp..
4	INFO	Henco, Inc
1,5,6	INFORMIX	Relational Database Syste
4	INFOTRIEVE	MRP Systems
1,5,6	INGRESS	Relational Technol. Inc.
1,5,6	INMAGIC	Warner-Edison Assoc Inc
1,5,6	INQUIRE	Infodata bSystems
4	INTAC	Ross Systems Inc
1,5	IV+V	Inst.f.Mach.Dokumentation

Table 4. Continued.

1	J-SPACE	Info. Acess Systems Inc.
2	K-MAN	Micro DB Systems Inc
1,5,6	LOGIX	Logical Software Inc
1,5,6	MARATHON	Relational DB Systems Inc
1,5	MDBS	Micro Data Base Systems
1.5.4	MINISIS	IDRC
1,5,6	MISTRESS	Rhodnius Inc.
1,5,6	ORACLE	Oracle Corp.
6	P-STAT	P-Stat Inc
1,5	POLARIS	Western Electric (AT&T)
4	PRODUCT 3	ELS Systems and Engineeri
1,5,6	RAPPORT	Logica, Inc.
2	RBASE 6000	Microrim Inc.
2	RTFILE	International Computing C
4	MESSIDOR	University of Paris
4	SDL	Source Data Systems Inc
1	SEED MICRO-SEED	Intern'l Database Systems
1,5	SEQUITUR	Pacific Software Mfg Co
1,5	SIR	Scient. Info. Retrieval I
1	SPIRES	Stanford University
4	SQL/DS	IBM Endicott Lab
1,6,5	STAIRS/VS	IBM
1,5	STAR	Cuuadra Associates Inc
4	STATUS	AERE
4	SYSTEM R	IBM
4	SYSTEM R*	IBM
4	TAGS	Database Systems Corp
1,5	TEXT RETR. SYST	MacKinney Systems Inc.
1,6	TOTAL	Cincom Systems, Inc.
3	TROLL	U.C.S.F.
4	UNIBASE	RLG Corp
1	UNIFY	North Am. Technology Inc
2	WATFILE	Waterloo Computing System
3	YARD	RLG Corp.

INFORMATION SOURCE LEGEND:

1 = vendor product description
2 = Stanford University [86]
3 = commUNIXations (UNIX newsletter), May 1983
4 = technical literature
5 = personal correspondence and communication
6 = commercial reports (DATADECISIONS, AUERBACH, DATAPRO)

G. FINDINGS

This section contains a discussion of the findings of the evaluation process described in the previous chapter.

1. "Universal" Database Software

It is useful to reiterate at this point the principal characteristics desired for the Egyptian STI Network's database software:

- Implementation, under UNIX-V operating system, on mini/microcomputers;
- Unattended operation and use in programmer-free environment;
- Near-universal capability as regards supported types of data (from integers to unstructured text);
- High-level, user-friendly command/query language;
- State-of-the-art processing capabilities transparent to the user;
- Mature product, supported by a responsible vendor committed to collaboration with Egypt.

The main conclusion emanating from the extensive analysis of available commercial software is that there appears to be no single software package that possesses all of the above characteristics. The reason for this somewhat disappointing conclusion is that the state-of-the-art of database software technology has not yet been able to bridge elegantly the disparities between information retrieval (text-oriented) and DBMS (data-oriented) systems. A few software packages appear to be closer to this goal than most; however, they failed the selection process for other critical reasons, as shown in Table 5.

Table 5. IR/DBMS SOFTWARE

DATABASE SOFTWARE	REASON FOR REJECTION
BASIS (Battelle Columbus Labs)	Not running under UNIX or in programmer-free environment
DB (Battelle Columbus Labs)	Not running under UNIX; not user friendly; untested
INQUIRE/IRMS (Infodata)	Cost ($185K per installation); too complex; not running in programmer-free environment; not available under UNIX; mainframe oriented
IV+V (IMD)	Not available until mid-1985; untested
MINISIS (IDRC)	Not running under UNIX; dedicated for H/P hardware; old design
STAIRS (IBM)	Not running under UNIX or in programmer-free environment; separate query languages used for data and text types

Of the above software, the relational IV+V package [37] merits attention in the future. It is being developed, with international usage in mind, jointly by the Institut für Machinelle Dokumentation in Graz (Austria) and the Deutsche Gesellschaft für Dokumentation in Frankfurt (West Germany), with some UNESCO involvement. The first release is slated for summer 1985; this will initially, however, be a single-user, micro-based package (see memorandum on IV+V status in Appendix I).

The relational MINISIS software, which also has UN roots and is used for bibliographic data systems in some 20 countries, has a very serious limitation in that it is written in assembly language and only for the Hewlett-Packard Series 3000 computers. Its adoption would force the Egyptian STI Network to standardize on this -- now relatively obsolete -- machine. In our conversations and correspondence with the International Development Research Centre in Ottawa, Canada, IDRC indicated that a) it had no plans to port MINISIS to UNIX, b) it cannot afford to rewrite MINISIS in a portable programming language, and c) India and one Latin American country planned to scale MINISIS down for a microcomputer (and presumably reprogram it in another language). Since MINISIS is nearly 10 years old, its program architecture and code are not state-of-the-art.

INQUIRE/IRMS is a fully developed nonrelational DBMS that handles both data types and information (text) retrieval. It runs only in IBM hardware and software environments at a purchase cost of $185,000 per installation.

DATA DECISIONS' system software user ratings (January 1984) show this package to be above average in vendor support and security provisions but below average in the following aspects: economy of resource utilization, ease of use, freedom from bugs, ability to expand processing volumes, and backup/recovery procedures.

Battelle Columbus Laboratories' BASIS is an online, inverted index data and text management system that runs on several mainframes and on DEC's VAX-11/70 minicomputers but not under UNIX. The software has evolved since 1973, and its purchase price (for a minicomputer) is about $65,000. The DATAPRO Report (December 1980) on BASIS concluded overall user satisfaction but voiced concern over the product's throughput and efficiency, and criticized the quality of documentation. Battelle Columbus Labs recently indicated an intention to port BASIS under the UNIX operating system, but no schedule has been made available to us.*

In summer 1984, Battelle Columbus Labs is introducing a new relational database manager called DM.* On the surface, DM has the characteristics of a "universal" DBMS. However, the software is programmer-oriented, and most applications will require programming; also, it will not run under UNIX in the foreseeable future. (In contrast, BASIS is a self-contained, modular system that requires no programming.)

Like BASIS, INQUIRE/IRMS and MINISIS, STAIRS is a mature, complex data/text management system dedicated to IBM mainframes. The system uses different query languages for data and text retrieval, respectively; the former is relational-like, the latter of the typical bibliographic kind. The software is available on lease only, for $750-$900 per month.

Theoretically, the alternative exists to build the appropriate software de novo. This has not been seriously considered because of the cost and effort involved. The IV+V software package had invested, by May 1982, 15 man-years in its development [64], and has already experienced a delay of 2-1/2 years. The cost of the DOMESTIC bibliographic software (a joint effort of West Germany and Israel) exceeded $1 million by 1983, long before it was completed.* The Egyptian STI Project has neither the financial resources nor the time necessary for such an effort, even though current software engineering techniques and tools might reduce the software development effort.

2. Relational DBMS Software

The relational model is widely viewed as the standard structure of future database management systems. The fifth-generation computer project of Japan, which intends to endow computers with natural-language capabilities and hence has to have provisions for sophisticated information query and search, selected the relational model as its standard. With this trend in mind, we have studied in detail a number of relational database managers that appeared, from the available information, to have some capabilities to handle text.

---

*Personal communication.

The second conclusion of this study is that none of the commercial relational DBMS systems have information retrieval facilities adequate to meet the respective requirements specified for the Egyptian STI system. At best, their existing text-related facilities are oriented toward word processing (WP) in office environments rather than toward information retrieval and management. Table 6 lists some of the negative characteristics of the relational DBMS software studied:

Table 6. Relational DBMS Software

RELATIONAL DBMS	REASON FOR REJECTION
DATABASE 2 (IBM)	IBM, non-UNIX environment only
DBMS (Prime)	PRIME, non-UNIX environment only
DBMS-II (DEC)	DEC, non-UNIX environment at this time
DG-DBMS (Data General)	DG, non-UNIX environment at this time
IMAGE (Hewlett-Packard)	H-P, non-UNIX environment at this time
*INGRES (Relat. Technol.)	Text data handling limited to WP at this time
*MISTRESS (Rhodinus, Inc.)	Cumbersome bibliographic IR applications
*ORACLE (Oracle corp.)	Poor text handling for IR applications
*POLARIS (AT&T)	Withdrawn from market in 1984
*SEQUITUR (Pacific Software)	Strong on WP functions but otherwise immature
SQL/DS (IBM)	IBM, non-UNIX environment only
TOTAL (CINCOM)	Too large and expensive; poor for documentation applications
*UNIFY (North Am. Technol.)	Poor text handling for IR applications

*Operates under UNIX

Table 6 shows that relational DBMS software designed by hardware vendors is intentionally restricted to their own machines. DBMS software developed by independent software houses does run almost invariably in the UNIX environment; unfortunately, its information-retrieval capabilities are inadequate. The inadequacy concerns the following important functions: free-text indexing and retrieval; severely limited automatic indexing (e.g., one keyword per attribute); the query language (usually SQL-type) is less user friendly than fill-in-the-blank prompts; word adjacency search capability non-existent; and full-text databases are not supported.

Of the UNIX-based relational DBMS systems, the user manuals of INGRES, MISTRESS and ORACLE were examined in depth, because of their leading commercial position. ORACLE has poor text handling facilities, and lacks the features listed above. INGRES has extended, in 1984, the maximum attribute size to 2000 characters but does not intend to support more sophisticated IR functions for at least a year or more.* MISTRESS features unlimited attribute length for textual data; and although it is used by a few customers for building bibliographic databases, the query language capabilities are limited. In sum, the relational DBMS design is progressing in its attention to text data types [6,13,19,50,53,59,88] but has neither implemented full IR system features as yet, nor does it seem to be tackling the issue of separate query languages for data and text types.

At least one university (Queen's University, Kingston, Ontario, Canada) has been building a relational information retrieval system for the past two years, but the effort is not close to completion.**

We have examined briefly the alternative of attempting to modify, under an agreement with its owners, an existing relational DBMS for information retrieval. The candidates informally approached (INGRES, MINISIS, MISTRESS and POLARIS) were not enthusiastic about such a proposition -- partly, we believe, because they do not (yet) consider the information retrieval market commercially attractive, partly because they have their own development plans and schedules.

---

*Personal communication, Prof. Stonebraker (Berkeley), April 1984.

**Personal communication, Prof. McLeod (Queen's U.), Oct. 1982, April 1984.

3. Information Retrieval Software

The progressive elimination of "universal" and relational database software had narrowed down the list of candidate systems to existing commercial, nontrivial information retrieval packages. Table 7 shows the names of systems examined and some of their critical characteristics.

Table 7. Information Retrieval Software

INFORMATION RETRIEVAL SOFTWARE*	SELECTED CHARACTERISTICS
BIBPRO-IV (Informatics)	Not running under UNIX
BRS/Search (BRS Inc.)	Mini/micro version of STAIRS-like mainframe software
FINDER (Aaron/Smith)	Full-text but single-user; incomplete; running on IBM-PC only
INMAGIC (Warner-Edison)	Not running under UNIX at this time
J-SPACE (Info. Access Sys.)	Not running under UNIX; English-language AI interface
SIR (Sci. Info. Retrieval)	Not running under UNIX
SPIRES (Stanford U.)	Not running under UNIX; complex; old
STAR (Cuadra Assoc.)	Dedicated, non-UNIX hardware
STATUS (ICL)	Not running under UNIX
TEXT RETRIEVAL (McKinney)	Similar to STAIRS; runs IBM only

*Other IR systems evaluated and eliminated earlier were BASIS, INQUIRY, MINISIS and STAIRS.

The third major finding of the study is that of these ten major information retrieval systems on the market, only one -- BRS/SEARCH -- was

designed to run and is fully implemented under UNIX. Vendors of several other IR software packages (INMAGIC, SIR, BASIS) have indicated intent to possibly port their software under UNIX; however, none were prepared to commit themselves to a delivery schedule.

Following the detailed examination of the BRS/SEARCH user's and administrator's manuals, a number of BRS clients were contacted, and a visit was arranged by a senior technical staff member of AT&T Technologies, Inc. and this consultant to BRS, Inc. headquarters in Latham, N.Y. on May 11, 1984. The full-day discussions concentrated on a) live demonstration of technical features, b) development plans and schedules, c) vendor support, and d) nature of possible contractual arrangements. As a result of these discussions, BRS, Inc. and AT&T Technologies agreed tentatively to port the BRS/SEARCH package to the AT&T 3b computer line, as a software product to be potentially marketed internationally by AT&T. The insights gained from this site visit are reflected in Appendix II.

H. RECOMMENDATIONS

1. Information Management Software

The Georgia Institute of Technology recommends that the Egyptian National STI Network purchase a license to use the BRS/SEARCH - Micro-Mini Version, Release 2.0 software as its Network-wide information database software. The software should be brought up on each computer purchased for the Network (3b5 and 3b2 computers). The vendor is Bibliographic Retrieval Services (BRS, Inc., Daniel Robinson, Vice President for Sales).

The detailed evaluation form for this software product is attached as Appendix II. (The form is the standard instrument recommended by the General Services Administration, Office of Software Development and Information Technology, to evaluate Federal procurement of software [24]). It is our belief that the BRS/SEARCH software is fully adequate for the information-system needs of the Egyptian STI Network, and that it is the most cost-effective package currently on the market.

To summarize the grounds for this recommendation for the benefit of the executive reader: BRS/SEARCH Micro-Mini version was introduced in late 1982 as a more advanced version of the mainframe system used by the company to support its international online database search system. The mainframe version is patterned after IBM's STAIRS software but it eliminates many of the DBMS components which give STAIRS its dual, complex character [78]. Since its introduction, the Micro-Mini Version has been implemented on some 30 computer installations in the U.S., and another 10 in Europe and the U.K.

BRS/SEARCH is written entirely in C, the native language of the UNIX operating system, and is optimized to interface with the UNIX system's structure, hierarchical file structure, as well as its utilities and calls. As a result, it is highly portable to other hardware environments using the UNIX executive software, and has been implemented on several different manufacturers' machines: DEC-11/34, DEC-11/70, DEC VAX 750, DEC VAX 780, FORTUNE, ONYX, and even the IBM-PC personal computer. The software is thus portable horizontally (among machines of different manufacturers) as well as vertically (from microcomputers to superminicomputers). It operates in a multiuser, multitasking, unattended mode, and features access permission techniques embedded in UNIX. System installation and maintenance are menu-driven.

From the functional viewpoint, BRS/SEARCH is a full-text management system featuring fully automatic free-text indexing, powerful search features geared toward unformatted natural language, a range of output print options (extendable by the formatting utilities of UNIX), and a set

of administrative utilities. The system is capable of, although not well suited to, handling data retrieval (as opposed to text retrieval); it also performs basic statistical operations on numeric data fields.

BRS/SEARCH is well documented in user and system administrator manuals. The source code is not available for purchase or licensing. The one-time fees for a license depend on the number of users (ports) on the central computer, and if undiscounted they vary from about $3,000 for a multiuser microcomputer to $30,000 for a large supermini machine.

BRS, Inc., originally a U.S. corporation, was acquired in 1982 by Indian Head, Inc., a New York information corporation. Indian Head, Inc., was bought subsequently by Thyssen-Bornemisza of the Netherlands Antilles, itself a subsidiary of Thyssen-Bornemisza of Monte Carlo, Monaco. (In addition to BRS, Inc., Thyssen-Bornemisza owns three other information-industry corporations in the U.S.)

Appendix II contains more details about the BRS/SEARCH software; Appendix III is a summary of interviews with three customers. The two manuals (System Administrator Guide and Users Guide) are provided under separate cover.

2. Data Management Software

The BRS/SEARCH should meet and exceed the information retrieval functions of the STI Network, and provide adequate software to handle any text-oriented databases. For predominantly numeric databases, however, information retrieval software is not ideal. The issue of database software for highly structured, tabular information remains open, therefore.

The Georgia Institute of Technology recommends that the issue of whether or not to procure a DBMS package (other than the BRS/SEARCH software) be deferred. There are several well-designed relational database managers on the market (INGRES, INFORMIX, MISTRESS, ORACLE) that can be purchased for the STI Network; however, it is our opinion that the purchase should be considered only after each of the sectoral information service nodes determines a) whether it intends to build or procure numeric databases, and b) if so, what types of data will be collected and stored. In our judgment, the four types of text-oriented, public databases discussed in Chapter D will initially consume considerable energies on the part of each sectoral ISN; in addition, many factual databases (e.g., a database of drugs, agricultural produce, manufactured goods, etc.) can very likely be accommodated by BRS/SEARCH. The possibility should also be investigated whether BRS/SEARCH might be extended to handle tabular (relational) data structures, inasmuch as UNIX embodies the basic tools for this purpose (the utilities PASTE, CUT, and JOIN). These reasons, the cost (on the order of $20 thousand per installation), and the complexity of having to master --and arabize--two complex software packages all argue in favor of deferring the purchase of a relational DBMS for a period of at least one year.

3. Administrative DBMS Software

There is one exception to the above recommendation. The STI Network management organization (housed at the ASRT), which does not provide public database services, should be equipped with a relational DBMS software. The functions of this software would be to a) collect administrative data from the sectoral nodes and provide management analyses, reports, plans, etc.; and b) to train the STI Network node personnel in the fundamentals and applications of relational DBMS software. The STI Network management organization's DBMS software would therefore be used for administrative, inhouse applications, not as a public resource.

Accordingly, the Georgia Institute of Technology recommends that a license be purchased for a relational DBMS software implemented on the 3b2 computer dedicated to Network management and administration. AT&T is about to announce (in fall 1984) support for several third-party DBMS systems of the INGRES caliber; the Georgia Institute of Technology will be pleased to evaluate the offerings at that time and submit an appropriate recommendation to Egypt. The anticipated purchase cost should be in the $3-5 thousand range.

I. REFERENCES

1. ASIDIC/EUSIDIC/ICSU-AB/NFAIS. Recommended Interchange Specifications for Computer Readable Bibliographic Data Bases. April 1978.

2. Bachman, C.W. Databases for Minis and Micros. Mini-Micro Systems, October 1981, pp. 149-154.

3. BCS Query Language Group. Query Languages - a Unified Approach. BCS Monographs in Informatics, Heyden Press, 1981.

4. Bestougeff, H. Designing the Database Management System. In: Database Management in Science and Technology, ed. by J.R. Rumble and V.E. Hampel, Elsevier Science Publishers B.V. (North-Holland), 1984. Pp. 119-166.

5. Biller, H. On the Architecture of a System Integrating Data Base Management and Information Retrieval. In: Research and Development in Information Retrieval, ed. by G. Salton and H.-J. Schneider. Springer-Verlag, New York, 1983. Pp. 80-97. (Lecture Notes in Computer Science Series, ed. by G. Goos and J. Hartmanis, Vol 146).

6. Blanning, R.W. Issues in the Design of Relational Model Management Systems. In: 1983 National Computer Conference. AFIPS Conference Proceedings, Vol. 52, ed. by A.N. Smith. AFIPS Press, Arlington, Virginia, 1983. Pp. 395-401.

7. Bole, L., ed. Natural Language Communication via Computers. Lecture Notes in Computer Science. Springer Verlag, Berlin, 1978.

8. Bookstein, A. Weighted Boolean Retrieval. Proceedings ACM-BCS Conference on Research and Development in Information Retrieval. In: Information Retrieval Research, ed. by R.N. Oddy, S.E. Robertson, C.J. van Rijsbergen, and P.W. Williams. Buttersworth, London, 1981.

9. Brooks, A.A. Data Types and Structures in Science and Technology. In: Database Management in Science and Technology, ed. by J.R. Rumble, Jr. and V.E. Hampel, Elsevier Science Publishers B.V. (North-Holland), 1984. Pp. 15-37.

10. Chapman, P. and P. Noerr. Blaise Cortex: A Microprocessor System for Libraries. Information Processing and Management, Vol. 19 (1983), no 1 , pp. 77-81.

11. Codd, E.F. A Relational Model of Data for Large Shared Data Banks. ACM Communications, Vol. 13, no. 6 (June 1970), pp. 377-387.

12. Coelho, H. Man-Machine Communication in Portuguese: A Friendly Library Service System. Information Processing and Management, Vol. 7

(1981), no. 2, pp. 161-181.

13. Crawford, R.G. The Relational Model in Information Retrieval. Journal of the American Society for Information Science, January 1981, pp. 40-64.

14. Cremers, W. Software Development with Supermicro and Unix. Online (Germany), Vol. 6 (July-August 1983). In German.

15. Croft, W.B. The Implementation of a Document Retrieval System. In: Research and Development in Information Retrieval, ed. by G. Salton and H.J. Schneider. Springer-Verlag, New York, 1983. Pp. 28-37. (Lecture Notes in Computer Science Series, ed. by G. Goos and J. Harmanis, Vol 146).

16. Deen, S.M. and P. Hammersley. Databases. Proceedings of the 1st British National Conference on Databases held at Jesus College, Cambridge, 13-14 July 1981. John Wiley and Sons, New York, 1981.

17. Deutsch, D.R. Progress Towards Database Management Standards. In: 1983 National Computer Conference, AFIPS Conference Proceedings, Vol. 52, ed. by A.N. Smith. AFIPS Press, Arlington, VA, 1983. Pp. 241-245.

18. Dierickxx, H. and A. Hopkinson. Reference Manual for Machine-Readable Bibliographic Descriptions. Unesco, Paris, 1982.

19. Dobosz, J. and B. Szymanski. An Implementation of Relational Interface to an Information Retrieval System. Information Systems, Vol. 6 (1981), no. 3, pp. 219-228.

20. Dodd, G.G. Elements of Data Management Systems. Computing Surveys, Vol. 1, no. 2 (June 1969), pp. 117-133.

21. Doszkocs, T.E. From Research to Application: The Cite Natural Language Information Retrieval System. In: Research and Development in Information Retrieval, ed. by G. Salton and H.-J. Schneider. Springer-Verlag, New York, 1983. Pp. 251-258. (Lecture Notes in Computer Science Series, ed. by G. Goos and J. Hartmanis, Vol 146).

22. Fayen, E.G. and M. Cochran. A New User Interface for the Dartmouth On-Line Catalog. Dartmouth University, Hanover, NH, n.d.

23. Fayen, E.G. Experimental On-Line Catalog for the Dartmouth College Libraries, Dartmouth University, Hanover, NH, April 1980.

24. Federal Conversion Support Center, Office of Software Development. The Software Improvement Process -- Its Phases and Tasks. Part 1. Office of Software Development, Falls Church, Virginia, July 1983. (Report OSD/FCSC-83/006).

25. Gault, F.D. Database Management Systems for Science and Technology,

ed. by J.R. Rumble, Jr. and V.E. Hampel, Elsevier Science Publishers B.V. (North-Holland), 1984. Pp. 39-73.

26. Gilbert, D., E. Parker, and L. Rosenthal. Microcomputers: A Review of Federal Agency Experiences. U.S. Department of Commerce, National Bureau of Standards, Washington, DC, 1983. (NBS Special Publication 500-102).

27. Godden, K. Categorizing Natural Language Queries for Intelligent Responses. Research Publication GMR-4639. Warren, MI, General Motors Research Laboratories, May 1984.

28. Gottschalk, C. Database Management Concerns at the International Level. In: Database Management in Science and Technology. Ed. by J.R. Rumble, Jr. and V.E. Hampel. Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1984. Pp. 241-263.

29. Gruber, W.H. and G. Sonnemann. Information Resource Management for Corporate Decision Support. In: 1983 Computer Conference. AFIPS Conference Proceedings, Vol. 52, ed. by A.N. Smith. AFIPS Press, Arlington, Virginia, 1983.

30. Harper, D.J. and C.J. van Rijsbergen. An Evaluation of Feedback in Document Retrieval Using Concurrence Data. Journal of Documentation, Vol. 34, no. 3 (September 1978), pp. 189-206.

31. Heaps, H.S. Information Retrieval—Computational and Theoretical Aspects. Academic Press, New York, 1978.

32. Hoey, P. O'N. Comparison of Cairs and Status Information Storage and Retrieval Software Over a 3 Month Period. Information Processing and Management, Vol. 18 (1982), no. 3, pp. 117-123 .

33. Hollaar, L.A., K.F. Smith, W.H. Chow, P.A. Emrath, and R.L. Haskin. Architecture and Operation of a Large, Full-Text Information-Retrieval System. In: Advanced Database Machine Architecture, ed. by D.K. Hsiao. Prentice-Hall, Englewood Cliffs, NJ, 1983. Pp. 256-299.

34. Horowitz, E. and S. Sahni. Fundamentals of Data Structure. Computer Science Press, Woodland Hills, California, 1976.

35. Houghton, R., Jr. Online Help Systems: A Conspectus. Communications of the ACM, Vol. 27 (February 1984), no. 2, pp. 126-133.

36. How to Evaluate and Select a Database Management System. Arlington Heights, Ill, International Software Enterprise, 1982.

37. Institut fuer Maschinelle Dokumentation. IV+V. Descriptive Brochure. Graz, Austria, 1984.

38. International Centre for Scientific and Technical Information. Communicative Format of Data Recording on Magnetic Tape.

International Exchange Format; MEKOF-2. International Centre for Scientific and Technical Information, Moscow, 1979.

39. International Federation of Library Associations and Institutions. UNIMARC: Universal MARC Format, 2nd revised edition. IFLA International Office for UBC, London, 1980.

40. International ISBN Agency. The ISBN System: Users' Manual. Second edition. The Agency, Berlin, 1978.

41. ISO/UNESCO. Information Transfer: ISO Standards Handbook 1. Second edition. ISO, Geneva; Unesco, Paris, 1982.

42. Joyce, J.D. and D.D. Warn. Command Use in a Relational Database System. In: 1983 National Computer Conference. AFIPS Conference Proceedings, Vol. 52, ed. by A.N. Smith. AFIPS Press, Arlington, VA, 1983. Pp. 247-253.

43. Khasanova, I.B. Criteria for Comparative Estimates of the Suitability of Database Management Systems. Naucho-Tekh. Inf. Ser. 2, no. 10 (1983). In Russian.

44. King, P.J.H. Database Management Systems -- A Technical Comparison. State of the Art Report. Pergamon Infotech, 1983.

45. Koll, M.B. WEIRD--An Approach to Concept-Based Information Retrieval, SIGIR Forum, Vol. 13, no. 4 (Spring 1979), pp. 32-50.

46. Koymen, K. An IS&R System Based on Turkish Language. Information Systems, Vol. 7 (1982), no. 3, pp. 201-207 .

47. Lacroix, M. and A. Pirotte. Comparison of Database Interfaces for Application Programming. Information Systems, Vol. 8, no. 3, pp. 217-229 (1983).

48. Lancaster, F.W. Information Retrieval Systems: Characteristics, Testing and Evaluation, 2nd edition. John Wiley and Sons, Inc., New York, 1979.

49. Lancaster, F.W. and L.C. Smith. Compatibility Issues Affecting Information Systems and Services. Unesco, Paris, 1983. PGI-83/WS/23.

50. Larson, J.A. Bridging the Gap Between Network and Relational Database Management Systems. Computer, September 1983, pp. 82-92.

51. Lee, D.T. Database-Oriented Decision Support Systems. In: 1983 National Computer Conference. AFIPS Conference Proceedings, Vol. 52, ed. by A.N. Smith. AFIPS Press, Arlington, Virginia, 1984. Pp. 453-465.

52. Lehot, P.G. and S.H. Kaiser. Network Database Systems. Architecture

Technology Corp., Minneapolis, Minnesota, 1984. (Report).

53. Macleod, I.A. A Data Base management System for Document Retrieval Applications. Information Systems, Vol 6 (1981), no. 2, pp. 131-137.

54. Maron, M.E. Probabilistic Approaches to the Document Retrieval Problem. In: Research and Development in Information Retrieval, ed. by G. Salton and H.-J. Schneider. Springer-Verlag, New York, 1983. Pp. 98-107. (Lecture Notes in Computer Science Series, ed. by G. Goos and J. Hartmanis, Vol. 146).

55. Maron, M.E., ed. Theory and Foundations of Information Retrieval. Drexel Library Quarterly, Vol. 14, no. 2 (April 1978).

56. Martin, J. Strategic Data-Planning Methodologies. Englewood Cliffs, NJ, Prentice-Hall, 1982.

57. McLeod, D. A Prescriptive Database Interface Methodology for End-Users. Information Systems. Vol. 7 (1982), no. 3, pp. 253-263.

58. Meadow, C. Applied Data Management. John Wiley and Sons, Inc., New York, 1976.

59. Moulinoux, C., J.C. Faure, W. Litwin. MESSIDOR: A Distributed Information Retrieval System. In: Research and Development in Information Retrieval, ed. by G. Salton and H.-J. Schneider. Springer-Verlag, New York, 1983. Pp. 51-61. (Lecture Notes in Computer Science Series, ed. by G. Goos and J. Hartmanis, Vol 146).

60. Morrissey, J. An Intelligent Terminal for Implementing Relevance Feedback on Large Operational Retrieval Systems. In: Salton, G. and H.-J. Schneider, Research and Development in Information Retrieval. Springer-Verlag, NY, 1983. Pp. 38-50. (Lecture Notes in Computer Science Series, ed. by G. Goos and J. Hartmanis, Vol. 146).

61. Negus, A.E. Standard Commands for Retrieval Systems. ECC, Bruxelles, May 1977. (Report).

62. North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development. Development and Use of Numerical and Factual Data Bases. AGARD Lecture Series No. 130. AGARD, Neuilly sur Seine, France, October 1983.

63. Page, J.R.U. Integrating Your System with Data Communication Networks and to Other Database Activities. In: Database Management in Science and Technology. Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1984.

64. Park, M.K. and R.E. Stearns, Jr. Evaluation of the IMD IV+V-System Software Package for Unisist Reference Manual Implementations. Unesco

Report. University of Georgia, Athens, May 1982.

65. Peat, L.R. Practical Guide to DBMS Selection. Walter de Gruyter, Berlin, 1982.

66. Pietilainen, P. Local Feedback and Intelligent Automatic Query Expansion. Information Processing and Management, Vol. 19, no. 1 (1983), pp. 51-58.

67. Project VIABLE Validates Relational DBMS Performance." Infosystems (Part 2). Vol. 31 (1984), p. 14.

68. Radecki, T. Incorporation of Relevance Feedback into Boolean Retrieval Systems. In: Research and Development in Information Retrieval, ed. by G. Salton and H.-J. Schneider. Springer-Verlag, New York, 1983. Pp. 133-150. (Lecture Notes in Computer Science Series, ed. by G. Goos and J. Hartmanis, Vol. 146).

69. Radhakrishnan, T., C. Grossner and M. Benoliel. Design of An Interactive Data Retrieval System for Casual Users. Information Processing and Management, Vol. 18, no. 1 (1982), pp. 23-32.

70. Robertson, S.E., M.E. Maron, W.S. Cooper. The Unified Probabilistic Model for IR. In: Research and Development in Information Retrieval, ed. by G. Salton and H.-J. Schneider. Springer-Verlag, New York, 1983. Pp. 108-117.

71. Roy, S. Buyer Beware - A Buyer's Guide to DBMS Software. DEC Professional, Vol. 2, no. 4 (July 1983), pp. 66-68/71-78.

72. Rumble, J., Jr. Data, Computers, and Database Management Systems. In Database Management in Science and Technology, ed. by J.R. Rumble, Jr. and V.E. Hampel. Elsevier Science Publishers B.V. (North-Holland), Pp. 1-13.

73. Rumble, J.R., Jr. and V.E. Hampel, eds. Database Management in Science and Technology. A CODATA Sourcebook on the Use of Computers in Data Activities. North Holland, Amsterdam, 1984.

74. Rumble, J.R., Jr. and N.L. Seiderman. Analysis and Display of Data in Science and Technology. In: Database Management in Science and Technology, ed. by Rumble, J.R., Jr. and V.E. Hampel. Elsevier Science Publishers B.V. (North-Holland), 1983. Pp. 75-91.

75. Russell, J. and D.E. Cullen. Implementing and Using the Database Management System. In: Database Management in Science and Technology, ed. by J.R. Rumble, Jr. and V.E. Hampel. Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1984. Pp. 197-209.

76. Salton, G. Automatic Information Retrieval. Computer, Vol. 13, no. 9

(September 1980), pp. 41-56.

77. Salton, G. Dynamic Information and Library Processing. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.

78. Salton, G. and M.J. McGill. Introduction to Modern Information Retrieval. McGraw Hill, New York, 1983.

79. Schmidt, J.W. and M. Brodie, eds. Relational Database Systems: Analysis and Comparison. Springer-Verlag, New York, 1983.

80. Shneiderman, B. Improving the Human Factors Aspect of Database Interactions. ACM Transactions on Database Systems, Vol. 3 (1978).

81. Shneiderman, B. Software Psychology: Human Factors in Computer and Information Systems. Winthrop Publishers, Cambridge, Massachusetts, 1980.

82. Simmons, P. and A. Hopkinson. CCF: The Common Communication Format. Unesco, Paris, 1984. (PGI-84/WS/4).

83. Small, D.W. and L.J. Weldon. An Experimental Comparison of Natural and Structured Query Languages. Human Factors, Vol. 25, no. 3 (June 1983), pp. 253-263.

84. Smith, F.J. Planning for Database Management in Science and Technology. In: Database Management in Science and Technology, ed. by J.R. Rumble, Jr. and V.E. Hampel, Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1984. Pp. 93-118.

85. Software Evaluation Guidelines. The Productivity Group, New York, 1982.

86. Stanford University. Microcomputer Data Base Management Systems: Their Selection and Evaluation. Stanford University, Palo Alto, California, 1984.

87. Stevens, G.C. User-Friendly Computer Systems? A Critical Examination of the Concept. Behav. Inf. Technol., Vol. 2, no. 1 (Jan.-March 1983), pp. 3-16.

88. Stonebraker, M., H. Stettner, N. Lynn, J. Kalash, and A. Guttman. Document Processing in a Relational Database System. ACM Trans. Office Inf. Syst., Vol. 1, no. 2 (April 1983), pp. 143-152.

89. Stonebraker, M. Operating System Support for Database Management. Communications of the ACM, Vol. 24, no. 7 (July 1981), pp. 412-418.

90. Tagg, R.M. Query Languages. In: Pergamon Infotech State of the Art Report on Programming Technology. Pergamon Infotech, Maidenhead,

1981.

91. Tagg, R.M. Query Languages for Some Current DBMS. In: Databases. Proceedings of the 1st British National Conference on Databases held at Jesus College, Cambridge, 13-14 July 1981. John Wiley and Sons, New York, 1981. Pp. 99-118.

92. Tague, J. and M. Nelson. Simulation of Bibliographic Retrieval Databases Using Hyperterms. In: Research and Development in Information Retrieval, ed. by G. Salton and H.-J. Schneider. Springer-Verlag, New York, 1983. Pp. 194-208. (Lecture Notes in Computer Science Series, ed. by G. Goos and J. Hartmanis, Vol. 146).

93. Teskey, F.N. Principles of Text Processing. Ellis Horwood, Sussex, England (Distributed by John Wiley), 1982.

94. Ulrich, W. Current Issues in Electronic Mail-Handling - Heralding a New Era. In: 1983 National Computer Conference, AFIPS Conference Proceedings, Vol. 52, ed. by A.N. Smith, AFIPS Press. Pp. 395-401.

95. United Nations Inter-Organization Board for Information Systems. Directory of United Nations Information Systems. 1. Information Systems and Databases. IOB, Geneva, 1980. 1980.

96. U.S. Department of Commerce. Federal Information Processing Standards Publications; Index. USDC, Washington, D.C., 1984.

97. van Rijsbergen, C.J. Information Retrieval. Butterworths, London, England, 1979.

98. Walker, D.E., H. Karlgren and M. Kay, eds. Natural Language in Information Science. FID Publication 551, Skriptor, Stockholm, 1977.

99. Waltz, D.L. An English Language Question Answering System for a Large Relational Data Base. Communications of the ACM, Vol. 21, no. 7 (July 1978), pp. 526-539.

100. Wiederhold, G. Database Design. McGraw Hill Book Co., New York, 1977.

101. Weiss, H.M. Which DBMS is Right for You? Mini-Micro Systems, October 1981.

102. Workowski, G. Relational DBMS Meets the Real World. Data Management, Vol. 21, no. 9 (September 1983), pp. 35-36/68.

103. What to Look for in Software Packages. DATAPRO Report AS60-300-001. DATAPRO Research Corp., Debran, J.J., July 1982.

104. Zloof, M.M. Query-by-Example: A Database Language. IBM Systems Journal, Vol. 16 (1977), no. 4, pp. 324-343.

APPENDIX I

STATUS OF IV+V SOFTWARE

GEORGIA INSTITUTE OF TECHNOLOGY
SCHOOL OF INFORMATION AND COMPUTER SCIENCE • ATLANTA, GEORGIA 30332 • (404) 894-3152

June 5, 1984

MEMORANDUM TO FILE

From: V. Slamecka VS

Subject: Status of the "IV+V" Database Software

The Institute for Machine Documentation, Graz, Austria (tel. 43-316-797730) is the site of the design and development of the relational database "IV+V", which is being considered by Georgia Tech as a candidate DBMS for the Egyptian national STI system. The following is a precis of my telephone conversation today with Prof. Werner Koch, Director of the Institute.

Status

Professor Koch stated that the approach to the software development is modular. Currently they have completed the input and the database processor modules. The output module, including a comprehensive report generator, is to be completed "within this year." The package is scheduled for release in mid-1985.

The software is implemented on different hardware processors via interpreters; currently interpreters exist for PDP-11/70 and some micros. UNIX environment would be eventually supported but the effort in this direction does not have highest priority.

Features

At present the software is intended to run in a single-user mode only, and is targeted at microcomputers and personal computers. Professor Koch stated that a multi-user version is part of their development plans but would not be available before spring 1985 (one potential customer is said to be interested in a multi-user version).

The package has no accounting software at the present time, but there are long-term plans to amend it with appropriate facilities.

As regards other parameters, the command language is based on SQL syntax but what the bibliographic database end user sees are fill-in-the-blank screens. The capability for designing customized screens is apparently a major feature of the system, and it allows individual sites to adjust the interface language to be compatible with, for example, the

A UNIT OF THE UNIVERSITY SYSTEM OF GEORGIA
AN EQUAL EDUCATION AND EMPLOYMENT OPPORTUNITY INSTITUTION

the Common Command Language of EURONET. The system user can specify the amount of automatic indexing on each attribute; thus, in principle, the system may support full-text storage and retrieval. Left- and right-hand truncation is also supported.

Availability

The Institute for Machine Documentation owns the IV+V copyright on a split basis with the German Society for Documentation. Marketing strategies have not been finalized. Professor Koch expects the cost of leasing the software to vary according to the package of components selected (which may include, for example, graphics capabilities). He anticipates the fees to run from $5K to $50K per installation.

The Institute is interested in testing the software in several experimental settings. Dr. Koch said he was active in Africa and Latin America. He did not believe that a complete product can be available to complement the STI Project schedule, although some modules are ready for testing.

Conclusions (by V.S.)

The IV+V software effort was conceived several years ago as a joint enterprise between the German Documentation Society and the Austrian Institute for Machine Documentation. In 1982, a team of three Unesco consultants reviewed the status of the project, and found that the plans were impressive, the approach sound, but very little existed in terms of actual accomplishment.[1] The initial release date was summer 1983; now it is said to be spring 1985. Despite the delays there is no reason to suspect that the effort is not highly professional and technically sound.

The major advantage of the package seems to be its universality to accommodate both textual (full-text and bibliographic) and non-textual data. The approach taken is to provide a screen-building interface that generates menus. The SQL language syntax used is transparent to the user of bibliographic databases; instead, he uses the same fill-in-the-blank approach as in the traditional retrieval systems such as Dialog or BRS. Presumably, SQL syntax is used with non-textual, non-bibliographic databases; if so, this means that the user has to learn two different command languages, depending on the nature of the database.

This design approach is similar to that of traditional relational database management systems such as INGRES, ORACLE, Sequitur, Mistress, Focus, and others, in that they provide facilities for the system administrator to design friendly interfaces for specific users, data types, and query languages. The major difference of IV+V is in its emphasis on handling bibliographic data with the relational model -- something that

none of the existing UNIX-based U.S. relational DBMS software emphasize (and, consequently, they are not well equipped to handle the storage and automatic indexing of full text).

From the viewpoint of its utility in Egypt the IV+V project has several current limitations. The major drawback is that it runs in single user mode; multiuser operation is not supported at the present time, and its implementation may be more time consuming than Dr. Koch anticipates. Other facilities essential to the Egyptian STI Network that are currently lacking are the output module (report generator), and an accounting module (allowing the system administrator to charge users for services). At this time, the software does not run in the UNIX environment.

By selecting to write the software in Pascal using the UCSD system as a development aid, developers have foregone the extensive capabilities of the UNIX programmer's workbench on which many of the U.S. relational DBMS developers rely; hence many of the prefabricated functions that already exist in UNIX may have had to be created for the IV+V system. Nevertheless, it is highly portable.

Although the leasing arrangements have not been firmed up, the quoted fee of $5K to $50K (depending on the configuration of software modules) per installation is high. As a single user system running on a $5,000 personal computer, the $5K fee seems over-priced. The $50,000 price is in the range of fully blown, large, multi-user relational database managers such as INGRES or ORACLE (which, in their micro version, sell for $1500 to $2000).

The IV+V system is thus likely to face more competition in the marketplace once the already existing database management systems are modified to handle text data more efficiently. (Some are moving in this direction already.) This notwithstanding, the backing of the United Nations, particularly Unesco, may provide the developers with a virtual marketing arm (and a source of funds) in developing countries.

In view of these considerations I recommend that Egypt remain au courant with this development at the Graz Institute. However, given the uncertainty as to when the IV+V software will become available as a tested, multi-user product running in the UNIX environment, it does not appear to be a viable DBMS candidate if the Egyptian project remains on its present schedule.

---

1. Park, M.K. and Stears, R.E., Jr, Evaluation of the IMD IV+V System Software Package for UNISIST Reference Manual Implementations. Paris, UNESCO, May 1982 (mimeo).

VS/al
cc: Mr. Bassit

APPENDIX II

BRS/SEARCH SOFTWARE PACKAGE EVALUATION FORM

EVALUATION CRITERIA FOR SOFTWARE PACKAGE

Package Name: BRS/SEARCH
Source/Vendor: BRS, Inc.

Evaluated by: V. Slamecka
Date Evaluated: May-June 1984

* Write remarks on a separate sheet of paper.

Evaluation Category and Criteria	% Weight Factor	% Rating	Remarks *
A. Functional/User Requirements			
1. Considering the overall consistency of the capabilities, do the basic functions of the package meet your requirements? List each function.			1
2. Considering the format and content of reports; characteristics, layouts, organization, and protection of files; error and audit routines; controls; and rerun provisions, do the detailed capabilities of the package meet your requirements? List each capability.			2
3. Will everyone who is affected by the package find that their needs have been satisfied? List those affected.			3
B. ADPE Configuration			
1. Considering your operating system, storage, tape and disk drives, printers card readers, etc., will the package run on your computer system as normally configured?			4
2. Does your system have the proper interfaces? If no, what is needed and what is the cost.			5

IV

EVALUATION CRITERIA FOR SOFTWARE PACKAGE

Package Name: _____

* Write remarks on a separate sheet of paper.

Evaluation Category and Criteria	% Weight Factor	% Rating	Remarks *
B. ADPE Configuration (continued)			
3. Does the package require additional hardware and/or software in order to run on your system? If yes, what is needed and what is the cost.			6
C. Software			
1. In what language is the package written?			7
2. Considering ANSI and FIP standards, are there any nonstandard coding conventions used in the package? List them.			8
3. Are your programmers familiar with the language?			9
D. Performance			
1. Is the performance of the package adequate for your needs?			10
2. Are other users who have a similar computer configuration satisfied with the package? List users' responses.			11
3. Did the package pass the benchmark and/or live test demonstration?			12
4. Is the package in an operational state or is it still being developed or modified?			13

EVALUATION CRITERIA FOR SOFTWARE PACKAGE

Package Name: _____

* Write remarks on a separate sheet of paper.

Evaluation Category and Criteria	% Weight Factor	% Rating	Remarks *
E. Attributes			
1. Is the package reliable? Explain.			14
2. Is the package efficient? Explain.			15
3. Is the package portable? Explain.			16
4. Is the package maintainable? Explain.			17
F. Package Content			
1. On what media will the package be delivered (e.g., magnetic tape or punch cards)?			18
2. Will the source code be delivered with the package?			19
3. Are package options available? List options and their cost.			20
G. Installation			
1. Is the package difficult to install? Modify? Explain.			21
2. Are any changes to the existing systems and procedures required? If yes, list changes.			22
3. Will any software or file conversion be necessary?			23

EVALUATION CRITERIA FOR SOFTWARE PACKAGE

Package Name: _____

* Write remarks on a separate sheet of paper.

Evaluation Category and Criteria	% Weight Factor	% Rating	Remarks *
G. Installation (continued)			
4. Will parallel operations be required? How much?			24
5. How much assistance, and at what cost, will the supplier furnish?			25
H. Documentation			
1. Is adequate documentation provided?			26
2. Does the documentation include a user's manual, narrative description, system and program flowcharts, source program listings, input/output layouts, and operating instructions?			27
3. Are the operating and/or maintenance instructions easy to follow?			28
I. Support			
1. How much support (e.g., technical or training) will the supplier furnish?			29
2. How much formal and on-the-job training is provided?			30
3. What are the qualifications of the suppliers's support personnel?			31
4. Will the supplier correct errors in the package after installation?			32
5. Will the supplier provide improved versions of the package?			33

EVALUATION CRITERIA FOR SOFTWARE PACKAGE

Package Name: _____

* Write remarks on a separate sheet of paper.

Evaluation Category and Criteria	% Weight Factor	% Rating	Remarks *
J. Costs			
1. Considering the direct and indirect costs of the package for items A - I, what is the total cost of the package?			34
2. What alternative financial arrangements are offered?			35
3. Will the package be bought or leased? List costs associated with each alternative.			36
4. Are there any special terms or conditions governing the package? List terms and conditions.			37
5. Are there any constraints placed upon the use or modification of the package? List constraints.			38
6. Is a discount offered for multiple installations of the package?			39
Totals	100%		

NOTES

1) BRS/SEARCH meets adequately the function of "information management" -- i.e., the storage and online retrieval of unformatted or partially formatted textual data in a multiuser environment. It meets all the mandatory requirements listed in Chapter E, STI Network Database Software Requirements. BRS/SEARCH is less adequate to handle highly structured, particularly numeric data that is best organized in tabular (relational) form.

From the viewpoint of the Egyptian environment, attention needs to be brought to several aspects of BRS/SEARCH, as follows:

a) BRS/SEARCH incorporates full-text, free-vocabulary indexing and searching. This is deemed crucial to the Egyptian environment for two reasons: i) it obviates the extraordinarily huge and difficult effort to build a single national thesaurus for science and technology (an effort not yet undertaken successfully by any country), and ii) it frees the users of the STI databases available in Egypt from having to have available to them copies of the many different thesauri and vocabulary lists used by foreign databases. BRS/SEARCH generates database indexes completely automatically from record files selectable by the database manager.

b) The above notwithstanding, some international databases used -- and contributed to -- by Egypt employ controlled vocabularies for human/indexing and retrieval. It should be noted that Version 2.0 of BRS/SEARCH does not have utilities for controlled vocabulary management and use (i.e., thesarus and other classification aids and authority lists). Thesauri and authority lists can, of course, be built and maintained as UNIX files, but the search routines of BRS/SEARCH do not access them interactively while the search is in progress. This limitation should not affect the STI Network nodes since all databases brought up on the ISN computers will use free vocabulary indexing. Foreign databases such as AGRIS which use controlled vocabulary will be reformatted for Egyptian use (the reformatting can, if desired, use only the controlled index terms, or preferably both controlled and free terms, in building the inverted index). Egyptian databases exported to organizations using controlled vocabularies must be indexed manually, and the terms included as a field ("paragraph") of the document record.

With respect to coded authority lists (e.g., corporate names) whose use reduces data entry time and error, the option exists to build them as UNIX files, and access them online during the document description process (using UNIX commands "grep" or similar), or make them available to indexers in printed form. The automatic substitution of the code with full entry can be

accomplished but a simple program may have to be written for this purpose.

c) Database updating is always from a file which is built offline. This feature is not by itself undesirable (the database entry function conserves database access); however, the correctness check of the updates is done by the program at the time the file is loaded, not at the time the file is built. This means that the data entry clerk must be able to use the UNIX editor (to build the file and to correct it if it will not load). The correctness verification routines are excellent, and programmable.

d) Once a database file has been structured it is not possible to add a new field to it. The only way to do so is by dumping the database into a file, editing it, and reloading it into a new file structure.

e) Adding new records ("documents") to BRS/SEARCH (and any other inverted-index) database tends to make the indexes less efficient (due to space management problems). Efficiency is regained through database reorganization. BRS/SEARCH accomplishes such reorganization automatically with a single command.

2) A BRS/SEARCH database is comprised of five basic files: the dictionary, inverted, text, text index, and information files. The location of the files is fully configurable. Each database text file consists of "documents" (records) subdivided into "paragraphs (fields), and "subparagraphs" (subfields). A database may contain up to 1 million documents; the overall size is limited by the system's disk storage. Each document may have a maximum of 65,000 paragraphs and subparagraphs (fields and subfields); the maximum length of a paragraph or a subparagraph is 255 sentences, where a sentence typically is a string of words terminated by a period, an exclamation point, or a question mark. Each sentence can have a maximum of 255 words, where a word is a string of alphanumeric characters surrounded by spaces or punctuation.

The configuration table for each database specifies the following main parameters: locations of the inverted files, text files, index files, dictionary file, information file, and database form file; location of intermediate work files needed during the load process; minimum privilege levels necessary to gain access to database and to execute the update, add and delete commands; the database password; and the name of the database manager having privileges to update or otherwise change the database.

The user configuration tables for each database contain file specifications, privileged specifications, and user specific variables. Among its contents are the user's privilege level, which can be a number between 0 and 255 (each user and each database have such privilege levels; a user is not allowed to sign onto a database

with a privilege level higher than his own); location of the user's permanent "save" directory and "temporary save" directory; location of the BRS/SEARCH message directory; location of the user's log trail file which contains every character keyed by the user and makes it possible to recreate a search session to locate user errors; and the specification of a particular database that the user always wishes to log into upon entering into BRS/SEARCH. (Some of these entries in the user configuration table are optional).

BRS/SEARCH has a multilayer security system based on UNIX capabilities. It consists of a login name, one or more password, a directory or file ownership, and directory and file "read," "write," and "execute" permissions for three categories of users: the owner, members of a definable group, and "others." Finally, as described above, a privilege level is assigned to every user and every database.

BRS/SEARCH offers a sophisticated, flexible, user-defined terminal driver system and a complete printer control system. Up to nine printers and one special batch printer can be used by the print queue; requests are processed by receipt number within priority order, and the system manager has authority to modify priorities.

The format and content of database output reports is flexible and definable (in Version 2.0) by the database manager or system user. The capabilities of the report generator are slanted toward textual rather than tabular data, but basic statistical functions can be performed on record fields containing numeric data.

3) As described in Chapter C, "Egyptian STI Network Software Schema," several categories of users are affected by the software: the system manager, who will find BRS/SEARCH to possess the tools necessary for administering the entire environment; ISN database searchers, who should find that the complement of search tools and aids adequately supports their search skills; the ISN data entry personnel, who will build files using the UNIX editor and load them into the database; and casual users with permissions to access and search databases who will be normally satisfied using a subset of the search commands repertory. A user guide is available online, as an option.

4) BRS/SEARCH will run unattended on both the 3b5 and 3b2 computers; no change in system configuration is required.

The size of BRS/SEARCH, including the user guide, menus, etc. is approximately 2Mb. The minimum CPU size for the full software package is 1Mb (both 3b2and 3b5 computers have a 2 Mb CPU). The memory required is twice the volume of the text files if no compression is used, or 1.3 with compression. The full package can be stripped down, if desired, to about 1.5 Mb.

5) BRS/SEARCH runs under the UNIX operating system furnished with the AT&T computers. BRS, Inc. will port the BRS/SEARCH software to the UNIX V version. No special interfaces for either of the computer

systems are required.

6) No additional hardware or software is required.

7) The package is written in C, the native language of UNIX.

8) Not applicable.

9) Several Egyptian members of the contractor's Cairo staff are partially familiar with the C programming language; the training project operated by the Al-Ahram Consortium includes a C programming course; and efforts are being made to urge computer science departments of Egyptian universities to introduce this language into their curricula.

10) BRS/SEARCH is a multi-user, sophisticated information management system. Its performance is not unrelated to the hardware facilities at each site, of course. One BRS, Inc. client, using a Vax 780 computer, reports that 20 to 30 simultaneous users are supported by this software. The AT&T 3b5 systems at the ISN nodes are configured for 12 user ports (8 hardwired; 4 remote). The configuration is expandable to accommodate larger numbers of users if necessary.

11) The Georgia Institute of Technology has selected three BRS, Inc. clients running the mini-micro version of BRS/SEARCH: Dartmouth University, AT&T Communications, and Kennametal, Inc. All interviews were positive as regards the BRS/SEARCH performance. The interview record is attached as Appendix III. European clients were not interviewed; their list is obtainable from Mr. Peter Gathin, Technical Director, BRS (Europe), London (telephone 44-1-580-5271).

12) No benchmark test was performed. A 1.5 hour demonstration has been arranged and performed by BRS, Inc. for the managers of information service nodes during their visit in the United States in December 1983 at the ASIS Conference exhibit in Washington, D.C. A one day technical presentation and extensive demonstration of the forthcoming 2.0 version of BRS/SEARCH was given to AT&T Technologies representative, Mr. Les Dickert, and Dr. V. Slamecka on May 11, 1984. The demonstration for the node managers ran on an Onyx 8002 system; the May 11 demonstrations were run on the Fortune computer and the PDP-11/34 machine. The performance of BRS/SEARCH was smooth in both demonstrations.

13) BRS/SEARCH has been fully operational since late 1982, and has been implemented at approximately 30 sites in the U.S. and 10 in the U.K. and Europe. However, BRS, Inc. continues to enhance the package continuously. The new version 2.0 is to be released in Summer 1984 (it enters beta testing on July 15). Earlier this year, BRS Inc. released the BRS/MAINT module which greatly facilitates the installation and administration of BRS/SEARCH by making it entirely menu-driven. The version 2.0 also contains an improved report

generator.

14) BRS/SEARCH is reported by the users interviewed to run in an entirely unattended mode, once installed. BRS, Inc. is well known in the information community for having an excellent telephone based support service.

15) The system design uses ordered sequential files and an inverted index. It is said [78] to be substantially more agile than STAIRS. As it is written in C, it relies on various UNIX features to optimize program execution. It is probable that the efficiency of BRS/SEARCH, when defined as response times in an interactive session, is likely to depend more on the disk access rate of the hardware system than on the internal efficiency of the BRS/SEARCH code.

16) BRS/SEARCH is fully portable in the UNIX environment, the standard environment of the STI Network.

17) BRS/SEARCH system maintenance is well-documented in the administrator guide. The guide describes procedures for adding, deleting, and changing users and databases; maintenance of the printer queue program; and backup considerations and procedures. The source code of BRS/SEARCH itself is not available, however.

18) BRS/SEARCH is provided in a variety of media. For the 3b5, it will be delivered on a standard 9-inch magnetic tape; for the 3b2, on 5-1/4 inch floppy disks.

19) No source code is available from BRS, Inc.

20) A complete set of options for the 2.0 version is not available but will be ascertained. Known options include the compression module, online user guide, various interface menus, and a text editor. The prices of these options have not been determined (see note 34 below).

21) With the BRS/MAINT front interface, installation is entirely menu-driven and highly user-friendly. The installation process consists of establishing the configuration directory, file location strategies, the system and database configuration tables, login options, and the user configuration tables; providing support for appropriate terminals; and installing the print module.

22) Not applicable.

23) No software or file conversion will be necessary.

24) Not applicable.

25) The supplier or AT&T are expected to provide a senior technical person on-site in Cairo during installation and to teach the BRS/SEARCH administrators course. The same or a different technical consultant will be brought to Egypt to hold a course for database searchers using

BRS/SEARCH software; subsequently such a course will be taught in Egypt by Egyptians trained by BRS, Inc. The costs of the installation support and training have not been determined but are expected to be within AID guidelines.

26) Yes.

27) The documentation includes the BRS Training Workbook, the User's Guide, and the System Administrator Guide. The latter contains UNIX manual type descriptions of BRS/SEARCH programs. Source program listings, input/output layouts, and program flowcharts are not available as part of the license. BRS, Inc., however, has indicated informally that it will work with Egypt to arabize BRS/SEARCH, so that Arabic databases and users speaking Arabic only can use the system. Such cooperation implies that Egyptian programmers involved in the arabization would be given access to the source code and program descriptions.

28) Yes.

29) See footnote 25 above.

30) See footnote 25 above.

31) The installation and training support personnel must be senior, highly experienced individuals.

32) Yes, or he will provide step-by-step instructions for errors that can be corrected locally and simply.

33) Yes.

34) The one-time license purchase fee for BRS/SEARCH applies to each installation, and is differentiated according to the number of ports on the machine. Currently, the fees run from about 5 to 30 thousand dollars, for installations with 4 to upward of 60 ports respectively. The license fees to the STI Network will be negotiated separately, inasmuch as AT&T is expected to become a discount-eligible OEM vendor for BRS/SEARCH and may set its own fee scale.

35) No alternative arrangements (e.g., annual leases) are currently offered by BRS, Inc. The alternative does exist, however, for the STI Network to purchase the software directly from BRS, Inc.; this appears less attractive than purchasing from AT&T since a separate maintenance agreement would have to be executed with BRS, Inc. The latter is not a large company, and it may not find it possible to devote adequate manpower resources to the STI Network installations.

36) See footnote 34.

37) Apparently none.

38) BRS, Inc. will be required to agree to such modifications of BRS/SEARCH as may be necessary to render the package bilingual, and to provide full cooperation and assistance in arabizing certain elements of the software. The constraints imposed on usage of BRS/SEARCH are those typical of copyrighted software and to licenses tied to single installations and CPUs.

39) See footnote 35 above.

APPENDIX III

TELEPHONE INTERVIEWS WITH BRS/SEARCH USERS

GEORGIA INSTITUTE OF TECHNOLOGY
SCHOOL OF INFORMATION AND COMPUTER SCIENCE • ATLANTA, GEORGIA 30332 • (404) 894-3152

July 2, 1984

MEMORANDUM

To: File

From: V. Slamecka

Subject: Telephone Interviews with BRS/SEARCH Users

BRS, Inc. (Mr. Dan Robinson, Vice President for Sales) provided a short list of BRS/SEARCH users in the U.S. From this list, three users were interviewed: one in a university, one in an industrial research laboratory, and one in an industrial library.

1. A telephone interview was conducted on June 28, 1984, with Miss Emma Vecht, AT&T Communications, New Jersey (telephone 201-993-5451). Miss Vecht is the database manager for the corporation.

AT&T Communications is running BRS/SEARCH on a VAX 780 system. The software has been running for the past two years, with one person in charge (Miss Vecht). There are several databases running at this installation under BRS/SEARCH, of which the largest one contains 27,000 records. At times, according to Miss Vecht, 20 to 30 concurrent users are online. Miss Vecht made the following points: BRS/SEARCH is the only bibliographic multiuser software system running in UNIX at this time. She considers it well-constructed from the systems viewpoint, and definitely user-oriented. The recently introduced BRS/MAINT is a menu-driven interface for system installation and administration by a novice user. The database is easy to change and edit online, using the Editor of UNIX. She considers the command structure (the fact that each command is preceded by two dots) to be somewhat of a drawback. (Note by VS: BRS/SEARCH found it necessary to differentiate its commands from those of UNIX, which are preceded by one dot.) For a database manager who knows UNIX, it takes no more than two days to learn the administration routines of BRS/SEARCH. Miss Vecht referred to the fact that BRS/SEARCH has "a lot of history," and that it has been a well-managed software effort.

2. A telephone interview has been conducted on June 27, with Miss Bonnie Motjar, Database Administrator at Kennametal, Inc., Pennsylvania (telephone 412-539-5000). This company is using BRS/SEARCH on a modified IBM personal computer. The IBM PC is running VENIX, a UNIX clone, in a multiuser, multasking mode; this has been accomplished by adding to the PC (which is a single user microcomputer) additional hardware marketed by Sritek, a Cleveland company. The Kennametal IBM PC also has a hard disk. This is said to be the

only BRS/SEARCH installation which runs on an IBM microcomputer in a UNIX environment. Miss Montjar, who is a systems programmer rather than a librarian, has been responsible for upgrading the IBM PC, for implementing the BRS/SEARCH package, and for designing and maintaining databases. She finds the BRS/SEARCH performance to be very good, and the "help" screens and BRS/MAINT to be "outstanding." She referred to "initial problems" with BRS/SEARCH, which had to do however with the unique IBM environment; namely, the BRS installation routine had to be custom modified because BRS, Inc. had not ported the software to the IBM PC at that time.

3. Telephone interview with Mrs. Emily Fayen, Dartmouth University Library, Dartmouth, New Hampshire (telephone 603-646-2574). Dartmouth University Library has been using UNIX for many years, first on a Terak microcomputer, currently on the VAX-11/750. BRS/SEARCH has been running first under UNIX Version 7, later under Berkeley 4.1, and currently under Berkeley 4.2 UNIX implementation. In moving from the Version 7 to the Version Berkeley 4.2, very few changes were needed, according to Mrs. Fayen. The largest database implemented has 365,000 records, each averaging 1,000 characters. With 8 to 9 simultaneous users being the peak load, the response on the VAX 11/750 is "quite good." Mrs. Fayen cautions, however, that periodic system updates, in which large amounts of data are entered in the database, are very consuming in both system resources and time. Although the update operation is entirely automatic, a monthly update will run for 15-20 hours. Dartmouth has found one "bug" in the software: if the system goes down, the console messages from the system make it appear as if the update was completed successfully. She stated that BRS was aware and fixing this program bug. Reformatting an existing file into the BRS/SEARCH database format is easy, and several front interfaces were available to reformat OCLC or RLIN records into local formats. In general, Dartmouth University Library was "very happy" with BRS/SEARCH.

4. BRS/SEARCH also runs on about ten installations in Europe. While no addresses were requested, BRS offered the name of Mr. Peter Gathin, Technical Director, BRS Europe, London, U.K., as a source of such data. His telephone number is 44-1-580-5271.

VS/al