INPUT

MANAGEMENT PLANNING PROGRAM IN INFORMATION SYSTEMS

VENDOR WATCH REPORT

MANAGING THE ACCEPTANCE OF OFFICE SYSTEMS

DECEMBER 1983

MANAGEMENT PLANNING PROGRAM IN INFORMATION SYSTEMS

OBJECTIVE: To provide managers of large computer and communications systems with timely and accurate information on developments that affect today's decisions and plans for the future.

DESCRIPTION: Clients of this program receive the following services each year:

- Impact/Planning Support Studies - In-depth reports dealing with the impact on users of projected managerial, personnel, and technological developments over the next five years. Studies include analyses and recommendations.
- Technology and Management Issue Briefs - Analyses of the probable moves of major computer/communications vendors in operating systems, data base/data communications ... value-added networks, personal compu...
- Residual Value ... residual values of major mainframe and ...
- Annual Plannin... ...asts of both short- and long-term pl... ...s by major industry classificatio... ...r.
- Conference ... convenient location in the fall qu...
- Executive B... ...usy executives to strategic iss...
- Inquiry Serv... ...nal research staff through tele... ...access to INPUT's library of re...

RESEARCH METH... ...ch in computers, communications, ar...

- Research to... ...ussions with client representati...
- Research fo... ...with users, vendors, universities, ...
- Conclusions ... ...dgment of INPUT's professional ...
- Professional ... 20 years of experience in dat... ...senior management positions with major vendors and users.

For further information on this report or program, please call or write:

INPUT
1943 Landings Drive
Mountain View, CA 94043
(415) 960-3990 Telex 171407

or

INPUT
Park 80 Plaza West One
Saddle Brook, NJ 07662
(201) 368-9471 Telex 134630

INPUT

INFORMATION SYSTEMS PROGRAM

VENDOR WATCH REPORT

MANAGING THE ACCEPTANCE OF OFFICE SYSTEMS

DECEMBER 1983

MANAGING THE ACCEPTANCE OF OFFICE SYSTEMS

CONTENTS

EXHIBITS

I INTRODUCTION

A. PURPOSE

- This report is part of INPUT's Information Systems Program (ISP). It identifies key factors and techniques for increasing acceptance of office systems.

- The report asks the following important questions:

  - What are the differences between sponsors, purchasers, and owners of office systems?

  - What are the differences between traditional systems and office systems?

  - What are the users' barriers to accepting office systems?

  - What are the risks associated with office systems?

  - How do office systems affect the information systems (IS) organization?

  - What type of support staff is required to improve acceptance of office systems?

- Which strategies should be developed to improve management and user acceptance of office systems?

B. SCOPE

- This report will focus on office systems that improve existing paper-based office procedures and computer-based and manual systems. It does not address technologies, such as optical disk and image processors, that have not yet been incorporated in most office systems environments. These technologies depend on central site information management strategies and do not require as much active involvement by end users.

- The following people should find this report pertinent:

  - IS managers.

  - IS planners.

  - End-user managers.

  - Senior corporate managers.

C. RELATED INPUT REPORTS

- Interested readers are referred to the following INPUT reports:

  - Impact of Office Systems on Productivity, November 1983.

- Establishes the framework for understanding the productivity problem and for evaluating office systems.

- The Opportunities of Fourth Generation Languages, September 1983.

  - Analyzes the extent to which fourth generation languages are used and how they fit into the information systems strategy.

- Organizing the Information Center, August 1983.

  - Discusses how to organize an information center, including chargeback methods.

- The Impact of the Office of the Future, December 1980.

  - Describes the expected effects of the "office of the future" on both the organization and the people within it.

- Managing the Integration of Office Automation in the EDP Environment, November 1980.

  - This report focuses on the tactical issues involved in managing the integration of office automation into the organization.

- Personal Computers Versus Word Processors: Resolving the Selection Dilemma, June 1983.

  - Compares and contrasts PC and WP roles inthe office environment for today and the future. It also includes a methodology to assist decisionmakers in making cost-effective selections that reflect each organization's unique environment.

- Selecting User Friendly Operating Systems for Personal Computers, June 1983.

  . This report establishes criteria and provides recommendations for selecting PC operating systems for different types of organizations.

D. REPORT ORGANIZATION

- The remainder of the report is organized as follows:

  - Chapter II is an executive summary.

  - Chapter III describes the importance of office systems to the organization.

  - Chapter IV identifies key factors to office systems acceptance from the perspective of management, organizations, and users.

  - Chapter V describes techniques for improving acceptance of office systems.

  - Chapter VI contains strategies for implementing successful office systems.

II EXECUTIVE SUMMARY

A. BARRIERS TO OFFICE SYSTEMS

- Office systems have evolved from standalone word processors to systems that integrate text, graphics, and computational functions for many users located in geographically remote locations.

- The cost and complexity of these systems coupled with the intangible nature of their benefits make justification difficult. Because of these factors management may perceive office systems as a high risk. There is also the belief by some managers that office systems are just a fad.

- Because of this complex situation, there are many barriers to acceptance of office systems. The difficulties are exacerbated by the diversity of office systems users, who range from CEOs using an executive information system to clerks using a word processor. Some of the barriers to acceptance include:

  - Rapid obsolescence of technology.

  - Fear of job displacement.

  - Inadequate support.

  - Distrust of IS.

- In many organizations, IS suffers a credibility problem with both users and managers. Users feel their requirements are ignored, and managers believe that information systems cost too much, do not satisfy needs, and have been implemented too late. Whether these perceptions of IS are valid is immaterial; their existence erects barriers to acceptance of office systems.

B. REMOVING THE BARRIERS

- IS must remove the barriers to accepting office systems, and their first step must be to improve their own image.

  - Their unresponsive image stems from the users' dependence on IS for all changes to their systems.

  - IS can change this by taking the role of facilitator for these systems, helping users to help themselves.

  - Users must perceive the systems as their own. IS must provide the support, allow the user to become self sufficient.

- Office systems are a corporationwide endeavor; their implementation should be directed by a task force comprised of managers, finance personnel, users, and IS.

  - Such an ecumenical group would nurture the users' ownership concept.

  - Including management and members of the finance department improves communications with the office systems' purchasers by improving their understanding of the benefits of the systems.

- Due to the potential high cost and the difficulty of measuring the benefits of office systems, pilot programs should be used.

  - The pilot should be selected carefully to assure success.

    - The pilot group should have a sufficient number of users (critical mass).

    - The user group should be visible to senior management.

    - The group should have an acute need that the system can satisfy (e.g., improved communications).

  - Beware of underestimating cost.

    - Don't have insufficient workstations or quality printers. Either could inhibit system use and shoot down the pilot.

    - Don't forget that a threshold exists at which additional processing power and storage will be required. At that point costs will increase greatly and present an unpleasant surprise if the threshold is not identified beforehand.

- Market office systems to the company. Produce newsletters that record the systems' successes, instruct the users on new functions, and give recognition to innovative users. Demonstrate the system to managers and present user testimonials on its worth to the organization. But don't oversell. False expectations can lead to dissatisfaction and rejection of the system.

- Finally, understand the users' fears and arrest them by demonstrating the benefits the system can deliver. Include users in the planning, development, implementation, and support stages. Giving users a sense of ownership is the key to removing the barriers to acceptance of office systems.

III THE IMPORTANCE OF OFFICE SYSTEMS

A. THE EVALUATION OF OFFICE SYSTEMS

- In recent years the office of the future and its miraculous effects upon American industry have been greatly hyped. Behind the futuristic claims, the mundane reality for most companies was an office system that consisted solely of word processing.

- The scene changed in 1982 when IBM announced its personal computer. An avalanche of personal computers reached offices across the nation, and they began to perform functions that were previously done by hand. Personal computers were legitimized.

- Those personal computers have evolved into office systems. They now demand communications not only from the mainframe computer for corporate data but from the other PCs within an office community. In fact, these PCs are actually intelligent workstations with the capability of becoming entry points to corporate, departmental, and office networks.

- Each company's position along this evolutionary network will be different. The salient facts about this evolution are that it is inexorable and that access to workstations will grow to include all strata of the corporation. Users' demands for these systems and for access to information will grow - at all levels.

- These systems are indeed office systems. The realization of their potential benefits to the company depends upon their acceptance by a vast and diverse user community. At this point, information systems (IS) organizations must take the lead to insure that users obtain systems that meet their needs. Understandable, helpful systems will lead to acceptance and progress.

B. RISKS OF NONACCEPTANCE

- Office systems are often discretionary. Electronic mail and filing systems require critical mass to be successful. If too few people use them, the systems will not produce promised benefits and may even be discontinued.

- Office systems are sometimes considered faddish. This image is enhanced by the large advertising expenditures of office automation vendors and the difficulty of justifying office systems.

  - INPUT's report, Methods of Cost Benefit Analysis for Office Systems, September 1983, described the complexity of justifying these systems.

  - Systems that are designed for professionals, such as decision support systems, are justified using references to potential improved decision making ability. Systems designed for clerical employees (word processors), however, can be justified by tangible methods, such as time savings and increased output.

- Due to the difficulty of financially justifying office systems used by professionals, a company's financial community may question their value.

- The skeptcism of the financial community implies a high risk of nonacceptance. Furthermore, the cost of providing access to office systems throughout the organization cannot be justified by traditional techniques. Managers who want these systems are under considerable pressure to somehow demonstrate their benefits to the company.

- The paradox is that the value of office systems is based on the exchange of information by many users. Critical mass is essential for success. But critical mass may require a large financial commitment.

- The alternative to installing integrated office systems is either to continue with manual systems or to use standalone personal computers.

  - Manual systems are collapsing under the weight of their own paper. The increasing requirements for easy access to perishable information in a usable format is making manual systems obsolete.

  - Personal computer systems are effective for personal computing. But the need to communicate and access corporate information is creating a demand for networking. In fact, networking is creating a de facto, but unplanned, office system. It is reactive, inefficient, and costly. The incremental expenditures of creating this system may exceed the cost of an integrated office system without acquiring all its capabilities.

- The risk of nonacceptance of integrated office systems may put companies into a reactive mode, a mode that has undermined traditional data processing for the past 20 years.

C. I.S. ROLE

- Whose system is it? Ownership of systems is a sticky problem. Especially since ownership implies responsibility.

- Traditional data processing systems are theoretically owned by the user but, in fact, are controlled by IS. This has led to many problems between IS and user organizations.

  - Traditional systems are developed and operated by IS for the user.

  - This makes users feel removed, and their requirements are not always met. This leads to a culpability chasm. IS believes it's the user's system and vice versa.

- To succeed, office systems must "belong" to the users. And IS must overcome its nonresponsive image.

  - The success of office systems is based on flexibility and ease of use. The user must be able to get and manipulate the information quickly and without using an intermediary - IS.

  - IS has the computing expertise to facilitate the use of these systems. Corporate data resides in IS-maintained systems. Access to this data and computer communications is in the purview of IS.

- IS is the logical organization to implement office systems but since these systems must be controlled by the user, the role of IS is changing. IS must become a facilitator.

  - IS must train, consult, and guide the user. IS personnel must remove obstacles that users perceive as inhibiting their operation of the system.

  - IS must change its reputation from that of an obstacle and to that of an asset to the user organization.

- Exhibit III-1 lists the functions IS must perform as part of its role as office systems facilitator.

- The next chapter discusses the key factors of office systems acceptance. Chapter V will then identify techniques that IS should employ to assure the highest probability of acceptance. But the key factor that permeates the acceptance is for IS to be and, most importantly, to be seen by the user as the facilitator for office systems.

EXHIBIT III-1

I.S.: O.S. FACILITATOR

Guide	Present the opportunities the systems provide to help users improve their job performance.
Train	Instruct in the capabilities and uses of the systems. Train users to train their peers.
Consult	Provide assistance on new uses of the systems.
Support	Provide detection and resolution assistance of hardware, software, and communication problems.
Interface	Be the liaison with other areas of IS to provide access to corporate data.

IV KEY FACTORS TO OFFICE SYSTEMS ACCEPTANCE

- To understand the factors of accepting office systems, two perspectives must be examined, the indirect and direct:

  - The indirect perspective is represented by the attitudes of management and by the philosophy of the organization.

  - The direct perspective involves users' attitudes toward information systems in general and toward office systems in particular.

  - This chapter examines both perspectives.

A. MANAGEMENT: DEVIL OR ANGEL?

- Broadway shows often look for a person to provide financial sponsorship. Because the show may not open without a major sponsor, Broadway coined the term "angel" for this very important person. In many instances, office systems require an "angel" to provide not only financial but, more importantly, political sponsorship of the system.

  - This person should be a manager, the more senior, the better. The manager's support can range from casual interest to hands-on use of the system.

- Management's interest is usually generated by need.

  - Executive information systems used by over 50 companies today provide senior executive access to strategic information from company and public data bases. These systems were sponsored by CEOs because of the need for immediate access to strategic information.

  - More commonly, managers sponsor systems that reduce workloads and improve communications. For example, a major international corporation was having problems meeting accounting deadlines for its international subsidiaries and in preparing periodic financial statements. The controller therefore sponsored an international electronic mail and filing system to reduce the communication and processing time.

- Sometimes managers believe that the company should have office systems and are willing to commit resources to assure that they are implemented. Although cost is important to them, the potential benefits, although intangible, are enough to justify the expense. Obviously, this "angel" must fly pretty high in the organization for such a justification to suffice.

• On the lowest end of the management sponsorship spectrum are the managers who are negative toward computer systems in general and office systems in particular. These managers think there is too much information available already. They believe that increasing the number of people that can produce, manipulate, and access data will only produce electronic garbage and reams of computer reports to be reviewed.

  - These managers believe that office systems are an expensive fad, and that the demand for these systems is artificially created by advertising agencies.

- Intangible justification is no justification to these managers. Unless a system can meet the same return on investment criteria as do other capital assets, it should not be installed.

- Although these managers may not be completely negative, the strategy must be to sell them the system. These managers are not angels.

• Most managers reside between the "angel" and the ultra conservative described above.

  - These managers must be convinced a system is a sound business decision.

  - Their acceptance or rejection of the system will be based on how well it meets the business objectives used to justify it.

  - Office systems' acceptance will depend on the same criteria, satisfying this business objective. Another important factor is user satisfaction. User attitudes will be discussed in section C of this chapter.

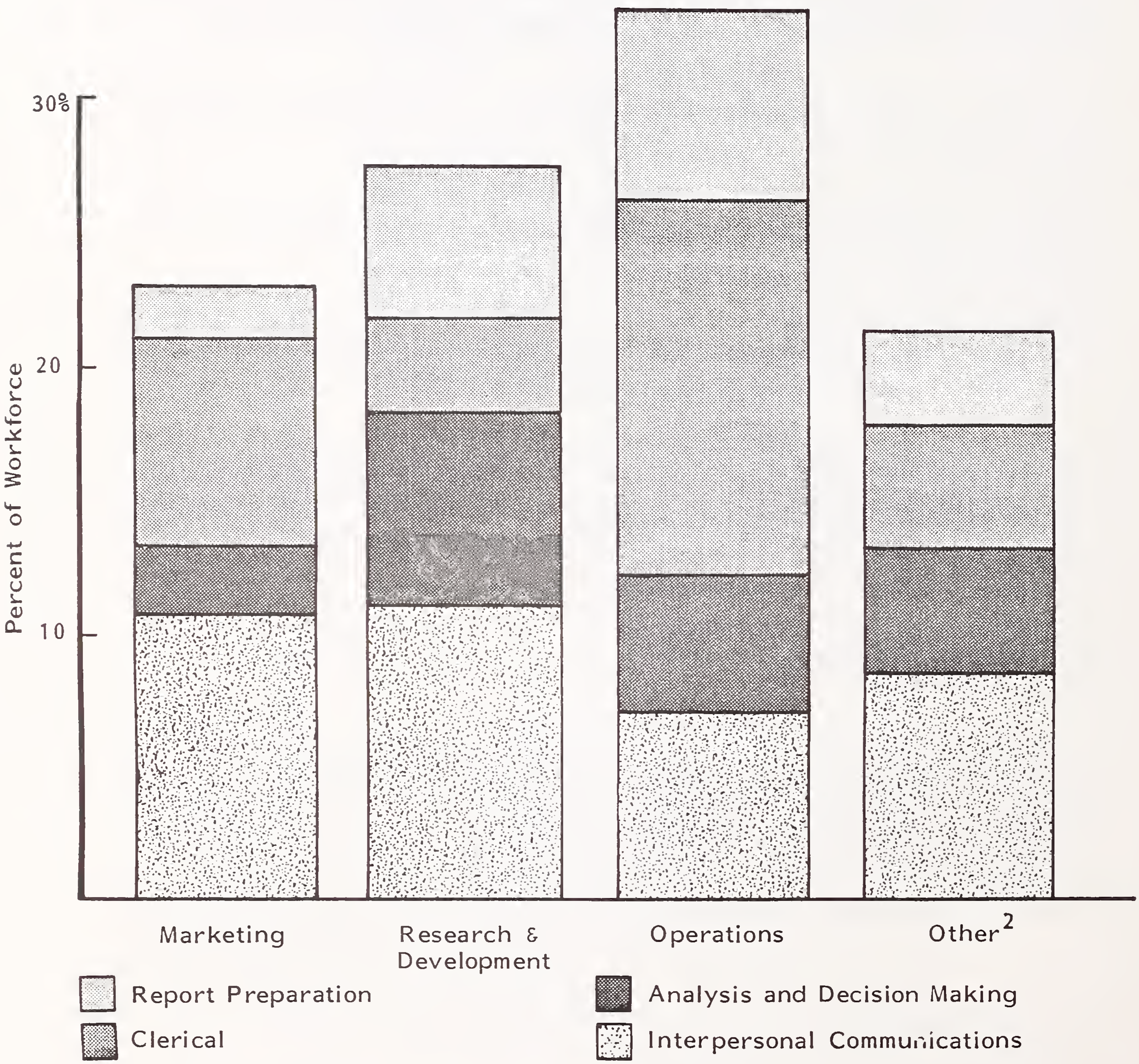
B. THE ORGANIZATION

• The demands for office systems vary by function and location, and satisfying them is a key factor in a system's acceptance.

  - Functional demands are related to the tasks performed. Each department has a different mix of functions that may be enhanced by office systems. Exhibit IV-1 reflects the distribution of these tasks by department.

EXHIBIT IV-1

OFFICE FUNCTIONS BY DEPARTMENT



[1] Includes: Information Systems and Engineering

[2] Includes: Personnel, Finance, Legal, and Senior Management

- The acceptance of office systems is founded on how well the system improves the performance of these functions.

- Systems that aid communications are more important to marketing than to operation departments. If a system causes problems initially in a critical area (communications in marketing, for example), it may be doomed even if the problems are resolved.

- Decentralized organizations present conflicting factors for accepting office systems.

  - Geographic dispersion leads to a strong desire for autonomy. Any vehicle that provides an opportunity for others (including headquarters) to meddle in their affairs will not be well received. Office systems may be perceived as a "spy system." Whether this is paranoid or not, this sub rosa feeling can sabotage an office system.

  - Geographic dispersion also demands better communications. This need is felt at both the remote facility and the headquarters. The ability to receive timely information is imperative for the success of most companies. Office systems can be the vehicle to satisfy this important need.

- Overriding the functional and geographical demands is the company's personality. If the company encourages open communication and close working relationships among departments, that will be the basis for accepting systems: systems that facilitate the free exchange of information. Companies that encourage autonomy are less concerned about corporationwide exchange of information and are more interested in each unit's performance. Systems designed to enhance the unit's performance will be valued. Their acceptance will be based on the unit's objectives and local concerns.

C. THE USER

- Ultimately, the user determines the system's success. In Chapter IV IS's role was described as that of facilitator. The user's role is that of owner of the system.

- The difference between office and traditional systems makes it difficult to define factors for user acceptance.

  - Traditional systems have a well-defined user community. There are standard reports that are delivered to fixed functional units. On-line access to these systems is also performed by designated functional units. These systems are rigid and the user, for the most part, must conform to the system.

  - Office systems are malleable. Their strength is that they allow users to transmit, access, manipulate, and report information as they choose. This strength can also be a weakness.

    - Lack of structure may make it difficult to diagnose problems.

    - The vast array of functions makes training difficult.

  - Office systems users can be found along the entire corporate hierarchy from clerical personnel using word processing systems, to the CEO using executive information systems.

- The range of users presents an array of potential obstacles for system acceptance.

  - Executives may be reluctant to use terminals because such use is inconsistent with their status.

- Cyberphobia, the fear of computers, may exist at any level of office system use.

    . Senior staff members may fear looking stupid to peers and subordinates.

    . Any staff member might fear causing a catostrophic error to the computer system, thereby disrupting the company's operation. (And then there is the fear that such a mistake would expose them to public ridicule.)

    . There is the fear of change, particularly among clerical workers who still fear losing their jobs to automation.

- The rapid changes in the office systems industry raise the spector of obsolescence. Systems may become obsolete before they have paid for themselves. This can postpone decisions to implement new office systems. There is also the fear that new systems will be incompatible with existing systems. They would require expensive modifications to either system in order for them to co-exist.

- Exhibit IV-2 lists the major barriers to acceptance of office systems.

• The next chapter discusses techniques for gaining acceptance of office systems and removing the barriers described above.

EXHIBIT IV-2

BARRIERS TO ACCEPTANCE OF OFFICE SYSTEMS

PERSPECTIVE	BARRIER
Management	• Image • Disenchantment with past management information systems • Belief that benefits cannot be translated to improved "bottom-line" results. • Rapid obsolescence and incompatible systems
Organization	• Destroys autonomy • Needs customization to specific needs • Inadequate support structure
End User	• Cyberphobia • Inadequate training • Job displacement • Distrust of IS • Fear of change

V TECHNIQUES FOR IMPROVING ACCEPTANCE

- The previous chapters described the office systems environment and the barriers to office systems acceptance. This chapter identifies techniques for improving that acceptance.

A. PILOTS - DON'T FLY BLIND

- Office systems are well suited for introduction using pilot programs.

  - Since the benefits of these systems are mostly intangible, a pilot program can demonstrate their benefits. The systems' usefulness can be proved or disproved in a real-life setting.

  - The financial risks can be mitigated by limiting the financial and staff resource commitments.

  - Office system software packages are available on a trial basis, and vendors encourage pilot programs to demonstrate their products. The theory is that once people are trained on the system, they will not want to give it up.

- Although pilot programs are good vehicles for implementing office systems, there are still obstacles that can cause them to crash:

- The wrong participants. Pilots require an enthusiastic group of participants who perceive the system as an aid to performing their job. People who are negative can shoot down a pilot before it is airborne.

- Too few participants. Communication is the glue that holds the various components of office systems together. Insufficient numbers of participants will not fully demonstrate the benefits of communication-based applications such as electronic mail. Sheer numbers alone, however, do not guarantee success. The users must comprise a group that interact while performing their normal job functions.

- Lack of visibility. If the pilot group is mired in the depths of the organization, its success will not be realized. Select a highly visible group in the organization. Their success will aid the expansion of the pilot. There is a risk in selecting a visible group, however: if the pilot crashes, the program may never fly again.

- False expectations. A pilot is actually a test program. The participants must understand that there will be problems but that people will be there to correct them. The pilot is not a production system; there may be turbulence. Most pilot participants understand this at the outset, but once the system takes off, they forget that this is a test and expect a smooth flight. An airpocket can be amplified because of false expectations.

- Unbalanced support. Training and rapid problem resolution is imperative to get the pilot off the ground. But too much support is also a problem. Office systems are user systems. They must become self-sufficient. Too much on-site support can be just as damaging as insufficient support. A proper balance must be struck to assure that the user can fly solo without crashing.

- The impediments to a successful pilot program can be removed by effective planning.

  - Users, managers, and IS must participate in this plan.

  - Status must be reviewed by participants, IS, and managers. Problems must be rectified quickly. Users must believe they have a voice in the changes to their system.

  - The pilot is a testing ground. Experiences must be documented. The system's use should be monitored not only to determine if it is to be used as expected but to see if any new, productive uses have been discovered.

  - Beware of tinkering. The system can become a toy. If users play with the system because it is fun, the result may be antiproductive. The pilot program provides the opportunity to develop procedures to guard against this problem. There is a very thin line between innovative productive uses of the system and unproductive tinkering.

- Exhibit V-1 summarizes impediments to successful office system pilot programs. Remedies are also listed for each impediment.

B. CUSTOMIZATION - THE KEY TO OWNERSHIP

- In Exhibit IV-1, the different mix of office functions was shown by department type. Each department considers different functions to be important. Similarly, each user may concentrate on a particular function.

  - Marketing departments use communications more than operations departments.

EXHIBIT V-1

PILOT IMPEDIMENTS AND REMEDIES

IMPEDIMENTS	REMEDIES
Wrong or Too Few Participants	Plan pilot with users and management to include groups that work together and frequently communicate.
Lack of Visibility	Work with senior management to select a group that has the most to gain. Make sure problems are quickly resolved, or the pilot may never fly again.
False Expectations	Develop a close working relationship with pilot group. Rectify problems quickly. Include users in an office systems steering committee system.
Unbalanced Support	Train users to be self-sufficient. Be responsive to user requests. Establish credibility with users and their managers.

- Market research analysts use computational tools more than most managers.

• Because of office system users' vast array of needs, a rigid system will not suffice. IS is also not structured to develop separate systems for each user group. Fortunately, office systems are general-purpose systems. They can be viewed as tools instead of programs.

  - The success of fourth generation languages is founded on simple syntax that nonprogrammers can use to program solutions.

  - Integrated software packages available on personal computers such as LOTUS 1-2-3 and VisiOn allow users to prepare reports using personal data bases and graphics.

• The general-purpose nature of office systems allows users to customize applications to fit their needs. Although these systems are user friendly, they cannot be merely turned over to the user.

  - Experience with personal computers has demonstrated that user friendly is truly a meaningless term. Many users have spent time trying to make these systems perform as advertised.

  - The need for data resident on different systems is a continual problem.

• The key to customization of office systems is to let users do it themselves, but with the guidance of IS.

  - IS can provide access to corporate information and the procedures to easily access this information.

- IS can provide general-purpose procedures to allow the user to easily move among functions. For example, using function keys to access electronic mail, file this information, retrieve related information, compose correspondence incorporating this information and mailing it to multiple recipients.

- IS should provide training so that users can customize features to satisfy their unique needs. If IS provides effective support, the user will gladly take ownership of the system. After all, they programmed it themselves.

C. COST - MANAGEMENT'S TOP CONCERN

- Cost is senior management's top concern in regard to information systems. To assuage it, vendors have developed integrated software packages that run on currently installed hardware.

  - IBM's Professional Office System (PROFS) runs on any 43XX or 30XX system that has a VM operating system installed.

  - DEC (All-In-One) and Data General (CEO) have integrated office system packages that run on their larger minicomputers.

- These packages may be leased or, in some cases, purchased for less than $50,000. This incremental cost, however, may be the tip of the iceberg.

  - When office systems become successful, demands for workstations, printers, disk, and ultimately processing power increase dramatically.

  - Of course, the vendor's strategy is to sell low-cost software in order to sell hardware.

- The vendor strategy is not necessarily detrimental to IS because it provides a relatively low-cost entry point for office systems. Also, pilots can be initiated with lower financial risk, therefore higher probability of success.

- Even though office systems pilots can be started on a nominal budget, don't over economize.

  - Office systems success is predicated on a critical mass of users. It must enhance office productivity, not reduce it.

    - An insufficient number of workstations will usually mean the system won't be used by a sufficient number of people.

    - An inadequate printer quality will lead to disuse of the system for text preparation.

  - The proper tools must be provided to the user, or the potential benefits of office systems will never be realized.

- Successful office systems have the potential of running rampant throughout the organization (remember the explosive growth of personal computers). Pilot programs may provide a good indicator of the benefits but a poor indicator of the cost of office systems.

  - A limited pilot may be started by adding a software package to a currently installed processor using workstations already in place. As the pilot grows, so do the hardware and software resource requirements. The system's success may lead to the purchase of additional hardware that was unforeseen during the pilot program.

  - Part of the planning process must include capacity planning. Volume constraints must be identified, and the cost of the system's expanding past that threshold must be identified and communicated to management.

- The cost of staffing resources is often overlooked.

  - IS must establish an effective support organization that is dedicated to office systems. This group must include educators and technicians.

  - Users will experience a learning curve when first using the system. During this start-up phase they will be less productive, which will mean hidden costs for the organization.

D. MARKETING - THE DELICATE BALANCE

- Office systems place IS in a new role, a system facilitator (see Chapter III). IS must provide a rich support organization to resolve problems, train, and assist users to become self-sufficient. But users must be convinced that office systems will benefit them. Management must be convinced that office systems are not a fad. And both groups must be convinced that IS is sympathetic to their needs.

- On the other hand, office systems should not be oversold. Unrealized expectations are the main cause of unaccepted systems. Remember, the first office system most people encounter is the telephone. New office systems are consciously or subconsciously compared to the telephone for function, ease of use, and reliability.

- This parodoxical situation must be addressed by IS. Office systems must be planned in order to be integrated into the company's business and information systems objectives. IS must take the lead in and effectively market office systems to the user, management, and the entire corporation.

- Marketing office systems requires knowing your customers and the role they play in the system life cycle.

  - Chapter IV described the factors for office systems acceptance. It looked at those factors from the viewpoints of management, the organization, and the user. Know the personality of your company and the individuals involved with the system. These people play one of the following roles:

    - Sponsor. This person, usually senior management, provides active support for the endeavor. The "angel" removes bureaucratic barriers impeding the system's success. Unfortunately, most companies don't have a sponsor for systems but they can be developed. Selling the concept of office systems should begin at the highest possible point in the company's hierarchy. The greatest benefit can be derived from office systems installed throughout the organization. This requires a senior executive's sponsorship. Angels must fly high.

    - Purchasers. These members of management are the toughest people to sell, because they pay for the system. These are the people most interested in tangible benefits and payback periods. The systems are in their budgets, and they are therefore responsible for justifying their expense. IS must make their job easier by providing the tools for justification. Provide techniques to identify all costs and benefits and provide guidelines for quantifying intangible benefits. For example, a multi-billion dollar conglomeration published a book on how to justify office systems for its managers. The techniques were proven by being used by managers throughout the organization and received the support of both the financial and audit departments.

- Users/owners. These are the ultimate accepters of the system. They must be convinced that the system will help them. It must be worth the risk of changing their work style and investing the time to learn the system.

- IS must understand users' needs and fears. Systems results must be realistically presented. IS's credibility problem in most organizations may be difficult to overcome, but a solution to business problems can transcend a poor image. If the system doesn't deliver the advertised benefits, the breech between IS and the user community may never be closed. The wary buyer of an office system is sometimes looking for an excuse to reject a system. Don't oversell and don't under support.

  - Honestly present the benefits and the cost to the user. Include all the costs identified in section C above, especially the hidden cost of lower productivity during the start-up phase.

  - The only way of closing the credibility gap between IS and the user is to provide effective support. Poorly organized, unresponsive support will destroy user confidence and undermine the system's success. Support personnel must be accessable to the user and must be responsive. Remember, users must be trained to be self-sufficient, or they will transfer ownership of the system to IS.

- Exhibit V-2 summarizes the roles of office systems users and identifies each group's primary strategic focus.

EXHIBIT V-2

ROLES IN OFFICE SYSTEMS' ACCEPTANCE

ROLE	POSITION	DESCRIPTION	FOCUS
Sponsor	Senior Management	Provide moral support to the system implementers. May eliminate bureaucratic bottlenecks for systems approval. Promote the system among their peers and encourage use of the system by subordinates.	Corporate
Purchaser	Management	Responsible for justifying and paying for the system. The key person for approving a new office system.	Department
Owner	User	Uses the system and is the key person involved with system acceptance.	Personal/ Work Groups

VI RECOMMENDATIONS

A. I.S. STRATEGIES

- Office systems success requires that the user take ownership of the system. This is not merely financial but also psychological.

  - Users must believe the system will help them do their jobs better. They must believe it's a personal system that they can customize.

  - Users must receive recognition for effective use of the system. Positive feedback will minimize their anxiety over the changing workstyle the system presents.

- Establish an office system task force comprised of representatives from management, finance, users, and IS. Attempt to have the task force chaired by a senior executive (a potential angel to office systems).

  - This task force will be responsible for planning office systems implementation throughout the organization.

  - It will be actively involved with pilot programs and will hold status meetings during development, installation, and postinstallation phases.

- This will be a political group rather than an IS group, which will be its major asset. Users will have access to this bipartisan group for their problems and suggestions. The stigma of being an IS system will be removed, and environment for user ownership will be established.

- IS should facilitate the use of the system.

  - Effective training and support is a prerequisite to successful office systems.

  - IS should establish an end-user training organization that understands the user's fears and arrests them through a combination of classroom, tutorial, and self-education programs.

  - User personnel should be trained to be trainers, to reinforce the concept of user ownership. The systems should be self-documenting with help-key functions contained in the system to answer common questions.

  - After implementation, IS should become training advisors and not primary trainers.

  - Part of the education process should include systems support features with the goal of making the user as self sufficient as possible. IS will still provide technical experts and support for problems and questions beyond the users' expertise. When IS is called, it must be responsive. Establish a hotline to answer questions, and be prepared to provide on-site support within a day of request for assistance.

- Select pilots with a high likelihood of success.

  - The group chosen for the pilot must have a need that is not being satisfied by its current systems and procedures.

- The group must be visible to senior management. This is a two-edged sword. If the pilot is unsuccessful, the future of office systems in the organization may be bleak. Even if problems are rectified, management support will be much more difficult to obtain.

- Realistic cost and benefit analysis must be performed.

  - Systems that do not deliver promised benefits and that exceed estimated costs are a thorn in management's side. In fact, this is the main cause of the IS credibility gap with senior management.

  - Remember, office systems costs resemble a step function. At a certain level of use, additional processors and storage media must be acquired.

  - Most of the nonclerical functions performed by office systems provide intangible benefits. Determine what the system purchasers require to justify a system, and design pilots to demonstrate those benefits (see INPUT's report Methods of Cost/Benefit Analysis for Office Systems, September 1983, for recommended justification procedures).

- Market office systems to the corporation.

  - Promote office systems successes throughout the organization. Establish a newsletter that describes the latest innovations accomplished by system users.

    - Provide recognition and encouragement to innovative users.

    - Promote system-assisted successes.

    - Share information among a potential, diverse user community.

- Remove the stigma of office systems being a fad.

  . Demonstrate productive uses of the system.

  . Implement simple features that have a potential for large benefit. Electronic mail, for example, can improve communication at relatively low cost in a regional, dispersed sales organization.

• Attack barriers to acceptance.

  - Exhibit IV-2 listed some of the barriers to acceptance of office systems. These should be attacked as part of IS office systems strategy.

  - Exhibit VI-I lists selected barriers to acceptance and strategies for removing them.

B. THE IMPACT OF OFFICE SYSTEMS

• Office systems can provide productivity improvements throughout the entire organization. This will only occur if office systems are planned.

  - Users must participate in all facets of office systems development.

  - Management support is essential.

  - Users must become self-sufficient.

• Office systems acceptance depends on user satisfaction. If users perceive the system to be their own, they will be less critical and more willing to solve problems.

EXHIBIT VI-1

REMOVING BARRIERS TO OFFICE SYSTEMS ACCEPTANCE

BARRIERS	REMOVAL STRATEGY
Management disenchantment with past systems User distrust of IS Management's concern for technical obsolescence	Establish joint management, user, and IS office system task force
Inadequate support, training, and structure Cyberphobia Users' fears of change and job displacement	IS should become an office systems facilitator, establishing an effective office systems support organization.
Management's belief that benefits cannot be translated into bottom-line results	Realistic cost/benefit analysis Marketing office systems
Customization	Flexible systems User ownership

- The alternative to planned office systems is independent systems that are hybrids of manual and personal computer systems. The cost of unplanned systems can be measured in increased computer expense (although it may be masked in user budgets) and lower productivity. Independently developed office systems will ultimately require interfaces with IS systems. The cost of these interfaces will be high, and IS personnel will be torn, trying to respond to numerous, unrelated requests. Many systems may be abandoned due to lack of support. User satisfaction will remain low and IS's credibility with users will not improve. In fact, IS may be blamed for not supporting these independent office systems.

- Office systems must be a corporate solution including management, users, and IS in its development. This team approach will improve the chance of acceptance and the realization of the vast potential benefits office systems can deliver.

MANAGEMENT PROGRAMS: Designed for clients with a continuing need for information on a range of subjects in a given area.

- Management Planning Program in Information Systems - Provides managers of large computer/communications facilities with timely and accurate information on developments that affect today's decisions and plans for the future.
- Management Planning Program for the Information Services Industry - Provides market forecasts and business information to information services companies to support planning and product decisions.
- Company Analysis and Monitoring Program for the Information Services Industry - Provides immediate access to detailed information on over 3,000 companies offering software, processing services, integrated systems, and professional services in the U.S. and Canada.
- Management Planning Program in Field Service - Provides senior field service managers in the U.S. and Europe with information and data to support their planning and operational decisions.
- On-Target Marketing - A practical, "how-to" methodology for more effective marketing problem solving and planning. Delivered to clients through workshops and/or consulting services.

MULTICLIENT STUDIES: Research shared by a group of sponsors on topics for which there is a need for in-depth, "one-time" information and analysis. A multiclient study typically has a budget of over $200,000, yet the cost to an individual client is usually less than $30,000. Recent studies specified by clients include:

- Selling Personal Computers to Large Corporations
- Improving the Productivity of Systems and Software Implementation
- User Communication Networks and Needs
- Financial Planning Systems Markets: The Next Five Years

CUSTOM STUDIES: Custom studies are sponsored by a single client on a proprietary basis and are used to answer specific questions or to address unique problems. Fees are based on the extent of the research work. Examples of recent assignments include:

- Organizing for Effective Software Development
- Corporate Plan for Utilizing CAD/CAM
- Annual ADAPSO Survey of the Computer Services Industry
- Analysis of Business Services for a Major Financial Institution
- Study of the Specialty Terminal Market
- Study of Disaster Recovery Services
- Analysis of Software Maintenance Issues
- Review of Software Product Market Opportunities
- Analysis of Network User Requirements

About INPUT

INPUT provides planning information, analysis, and recommendations to managers and executives in the information processing industries. Through market research, technology forecasting, and competitive analysis, INPUT supports client management in making informed decisions. Continuing services are provided to users and vendors of computers, communications, and office products and services.

The company carries out continuous and in-depth research. Working closely with clients on important issues, INPUT's staff members analyze and interpret the research data, then develop recommendations and innovative ideas to meet clients' needs. Clients receive reports, presentations, access to data on which analyses are based, and continuous consulting.

Many of INPUT's professional staff members have nearly 20 years' experience in their areas of specialization. Most have held senior management positions in operations, marketing, or planning. This expertise enables INPUT to supply practical solutions to complex business problems.

Formed in 1974, INPUT has become a leading international planning services firm. Clients include over 100 of the world's largest and most technically advanced companies.

OFFICES

Headquarters
1943 Landings Drive
Mountain View, CA 94043
(415) 960-3990
Telex 171407

Detroit
220 E. Huron
Suite 209
Ann Arbor, MI 48104
(313) 971-0667

New York
Park 80 Plaza West-1
Saddle Brook, NJ 07662
(201) 368-9471
Telex 134630

United Kingdom
INPUT, Ltd.
Airwork House
35 Piccadilly
London, W1V 9PB
England
01-439-8985
Telex 23116

AFFILIATES

France
La Nacelle
2, rue Campagne Première
75014 Paris
322 56 46
Telex 220064 X5533

Italy
PGP Sistema SRL
20127 Milano
Via Soperga 36
Italy
Milan 284-2850
Telex 310352

Japan
Overseas Data Service Company, Ltd.
Shugetsu Building
No 12 - 7 Kita Aoyama
3-Chome Minato-ku
Tokyo, 107
Japan
(03) 400-7090
Telex 26487

Sweden
Athena Konsult
P.O. Persson & Co AB
Box 22114
S-104 22 Stockholm
Sweden
08-52 07 20
Telex 17041

INPUT Planning Services for Management

INPUT

MANAGEMENT PLANNING PROGRAM IN INFORMATION SYSTEMS

INFORMATION SYSTEMS

ISSUE REPORT

RELATIONAL DATA BASE DEVELOPMENTS

AUGUST 1983

MANAGEMENT PLANNING PROGRAM IN INFORMATION SYSTEMS

OBJECTIVE: To provide managers of large computer and communications systems with timely and accurate information on developments that affect today's decisions and plans for the future.

DESCRIPTION: Clients of this program receive the following services each year:

- Impact/Planning Support Studies - In-depth reports dealing with the impact on users of projected managerial, personnel, and technological developments over the next five years. Studies include analyses and recommendations.

- Technology an… the probable moves of major compute… ystems, data base/data communicatio… value-added networks, personal compu…

- Residual Value… esidual values of major mainframe and…

- Annual Planni… of both short- and long-term pla… major industry classification.

- Conference - … nvenient location in the fall qua…

- Executive Bul… y executives to strategic issue…

- Inquiry Servic… l research staff through teleph… ccess to INPUT's library of rese…

RESEARCH METHOD… in computers, communications, and…

- Research topi… ions with client representative…

- Research for t… th users, vendors, universities, in…

- Conclusions de… ment of INPUT's professional st…

- Professional staff supporting this program average nearly 20 years of experience in data processing and communications, including senior management positions with major vendors and users.

For further information on this report or program, please call or write:

INPUT
1943 Landings Drive
Mountain View, CA 94043
(415) 960-3990 Telex 171407

or

INPUT
Park 80 Plaza West One
Saddle Brook, NJ 07662
(201) 368-9471 Telex 134630

INPUT

INFORMATION SYSTEMS PROGRAM

INFORMATION SYSTEMS

ISSUE REPORT

RELATIONAL DATA BASE
DEVELOPMENTS

AUGUST 1983

RELATIONAL DATA BASE DEVELOPMENTS

CONTENTS

EXHIBITS

I INTRODUCTION

- Data Base Management Systems (DBMSs) were conceived approximately 20 years ago and have engendered continuing technical, political, and emotional controversy. Today, Relational Data Base Systems (RDBSs) serve as a focal point for this ongoing controversy. The purpose of this report is to:

  - Define relational data base systems in understandable terms.

  - Evaluate the advantages and limitations of RDBSs (as defined in this report).

  - Project future directions in both software and hardware implementations of RDBSs.

  - Provide rough guidelines for user selection and application of RDBSs.

- For this report on relational data base systems, product evaluation was ruled out since literally hundreds of "relational-like" and "semi-relational" systems have been defined and announced; as recently as November 1981, the originator of the relational model (E.F. Codd of IBM) stated that he was not aware of any "fully relational" systems having been developed. Therefore, even user experience with currently available systems would not accurately reflect either the potential or limitations of the relational model.

- The research was structured with certain specific goals, which permitted a narrowing of the available information sources. A limited telephone survey of current INPUT clients revealed that their primary interests were 1) IBM direction in relational data bases, and 2) the applicability of the relational model to large data bases. This survey helped establish a framework for emphasis.

- For the above reasons, special attention was given to IBM publications concerning RDBSs, reports authored by IBM authors, and the analysis of IBM's apparent DBMS strategy. General technological trends as they relate to RDBSs are interpreted primarily in terms of impact on IBM strategy. Telephone interviews were used to verify research results, and to refine the conclusions reached. Previously published INPUT reports relating to this subject are listed in the Appendix.

- As this report neared completion, IBM announced Database 2 (DB2), which is a relational data base management system for MVS/XA and MVS/370 architectures. This gave INPUT an opportunity to briefly analyze DB2 and place it into perspective in terms of the IBM strategy.

II EXECUTIVE SUMMARY

A. INTRODUCTION

- Relational data base systems are currently being promoted as the solution to two serious problems - ready access to data bases by end users and improving productivity in the systems development process. While these problems have in some measure been caused by past data base management systems (DMBSs), the relational systems do have the potential for solving both problems at least partially.

B. DEFINITION OF RELATIONAL DATA BASE SYSTEMS (RDBSs)

- There are three fundamental components of any data model:

  - A defined set of data structure types.

  - A collection of operators or rules of inference to derive, modify, or retrieve data from the defined data structure types.

  - A general set of integrity rules that implicitly and explicitly define a set of consistent data base states or changes of state or both.

- In the simplest possible terms the relational model can be described as follows:

  - The relational data structure is tabular and consists of rows and columns.

  - The basic operators (relational algebra) that operate on rows and/or columns of the data structure are:

    . <u>Select</u> takes one relation (table) and produces a new relation consisting of rows of the first.

    . <u>Project</u> transforms one relation (table) into a new one consisting of selected columns of the first.

    . <u>Join</u> takes two relations (tables) as operands and produces a third table. The third table consists of the rows of the first concatenated with the rows of the second and is constructed by matching values in the columns of the first two tables.

    . The primary purpose of these operators is loop avoidance, and the basic operators are not intended to be a standard language.

  - The general integrity rules consist of five principal normal forms. These forms are useful as guidelines for data base design regardless of whether an RDBS is employed. (The normal forms are briefly described in the body of this report; the guide to the five normal forms that are cited above is recommended.)

C. ADVANTAGES AND DISADVANTAGES OF RDBSs

- The primary advantages normally associated with RDBSs are flexibility and ease of use. These advantages apply to both professional systems developers and end users; they are considered to be of special importance because of current trends towrads prototyping, information centers, and decision support systems.

- An advantage which is not so frequently mentioned, but which INPUT considers to be especially important, is the communicability of the simple relational structure (tables) among users, programmers, and data base administrators - especially in the current environment where end-user development is being encouraged. In fact, bridging the communications gap between the corporate data base designers and end users could be the most important contribution of the RDBS.

- The primary disadvantage of relational systems has been and continues to be performance. There are costs associated with flexibility and simplicity even when an RDBS is used properly. When misused an RDBS can result in a prohibitively expensive system.

D. FUTURE SOFTWARE AND HARDWARE DIRECTIONS

- RDBSs, for the reasons mentioned above, are especially well suited for end-user development, and this means the trend towards personal computer implementation of RDBSs will continue.

- During the preparation of this report, IBM announced that it will make available an RDBS (Data Base 2 (DB2)) that is compatible with its mainline operating systems (MVS and VM/XA). A clear picture of IBM's future software direction is provided by their choice of an interface (DXT) that extracts from both large production files and archival data.

- The ability to build relational structures from existing data bases and files will result in rapidly expanding use of DB2 on large IBM mainframes. Inevitably, this will place substantial burdens on current IBM hardware/ software systems because of the RDBS performance problems previously mentioned.

- The solution to these performance problems will lead logically to new hardware systems that are more appropriate for the implementation of RDBSs. The packaging of the new hardware may vary (it may be in the form of mainframe architecture, intelligent controllers, or data base machines), but the features can be seen in current data base machines. INPUT predicts IBM will find it necessary to seek a hardware solution to DB2 performance problems shortly after the system becomes available in 1984.

E. ANALYSIS OF RECENT AND PROJECTED RDBS DEVELOPMENTS

- The relational model has been debated for over 10 years without its being clearly defined or understood. This situation is now being alleviated by the publication of a large body of meaningful and understandable publications.

  - The relational model is largely the work of E.F. Codd (IBM Fellow of the San Jose Research Laboratory). His definitions, as contained in the 1981 ACM Turing Award Lecture ("Relational Database: A Practical Foundation for Productivity," Communications of the ACM, February 1982), are proposed as standards against which to measure relational data base systems implementations.

  - Simple definitions of the relational normal forms have been extracted from "A Simple Guide to Five Normal Forms in Relational Data Base Theory" (Communications of the ACM, February 1983).

- Performance evaluation information on relational systems has been extracted from "A History and Evaluation of System R" (Communications of the ACM, October 1981) and "System R: An Architectural Overview" (IBM Systems Journal, Volume 20, Number One, 1981).

- One of the primary recommendations of this report is that the relational model and its hardware/software performance implications should be thoroughly understood by those planning to install an RDBS (expecially for use with large data bases in a complex operating environment). This report could not (and does not pretend to) provide all of the information necessary for such understanding, but it does provide conclusions that should prompt additional research.

- Support of the "relational solutions" was materially enhanced recently when IBM announced DB2 under MVS/XA. However, it is believed that there are two inherent dangers in unqualified acceptance of the solution:

    - The expense and complexity of maintaining separate data bases for production and planning.

    - The potential performance impact on the large host processors, which may not only be expensive but unworkable.

- It is the very flexibility and ease of use of RDBSs that insures that multiple overlapping data bases will occur and that there will be inherent performance problems associated with implementation of the relational model.

    - Duplicate data bases will occur because DB2 and comparable systems encourage the extraction of data from existing data bases and files but do not provide for the elimination of those files.

- The conversion of large production data bases to the relational model will be both difficult and expensive.

- Updating and synchronization between the duplicate data bases is difficult and clumsy - assuring long life for existing production data bases.

- At another level, the problem is compounded by relational data base systems on both minicomputers and personal computers.

- Even if production data bases could be replaced, the ensuing impact on systems performance would probably be intolerable.

  - The experience with System R (on relatively small data bases with limited users) indicates that performance is a key consideration and that all of the problems have not been solved.

  - Regardless of how much improvement can be made by clever implementations of relational systems, a price must be paid for the relational model's benefits. While RDBS performance may be acceptable, it is not likely ever to be as good as systems with indexed or linked navigation through the data structures.

- The performance problems lead us to believe that data base machines will become increasingly appealing for the implementation of relational systems. Such systems hold promise for both architectural and geographical distribution of processing (and therefore data base).

  - Backend data base machines (DBMs) are already available; it is projected that IBM will provide, shortly after DB2 is delivered in 1984, a hardware assist for data base management on the backend. However, its actual hardware may take the form of a more intelligent controller rather than a full-function DBM.

- It is doubtful that IBM will substitute DBMs for mainframes in a distributed data base environment. At present, this is probably just as well since problems of interfacing between distributed data processing (DDP) and office systems remain to be resolved.

- As distributed processing and data bases penetrate the office environment there will be a need to integrate the management of information - data, text, images, and voice. It is intuitively felt that RDBSs and DBMs will contribute substantially to the solution of some of these problems and that the problems of data and information management should receive emphasis in artificial intelligence (AI) and advanced computer systems.

  - The tendency of AI to emphasize the acquisition of "knowledge" has frequently ignored the unsolved problems of knowledge storage and access to data and information.

  - The trend to supercomputers that overpower the problems associated with knowledge-based systems (such as the Japanese fifth generation) may not bridge the gap between data base management and the requirements for information management in the office.

  - The relational model, trends to new architectures in DBMs, and associative memories seem to hold the most promise of bridging this gap.

F. GUIDELINES FOR SELECTION AND USE OF RDBSs

- Without careful analysis and planning, current relational data base technology (both hardware and software) cannot be relied upon to solve even today's problems. Specifically, RDBSs should be properly applied and not viewed as a universal solution to an extremely complex problem set.

- Information Systems (IS) management is urged to:

  - Apply Codd's definitions in the selection of RDBSs to assure at least minimum functional (and structural) capability.

  - Become familiar with the relational normal forms and apply them in data base design regardless of the data model employed.

  - Standardize the use of personal computer RDBSs.

  - Either monitor and control the use of relational data bases on PCs, <u>or</u> be prepared for these data base to expand onto mainframes.

  - Monitor and control the extraction of relational planning data bases from current data bases or files - <u>or</u> be prepared for these planning data bases to evolve toward large-production data bases with resulting performance problems.

III BACKGROUND AND STATUS OF RDBSs

A. BACKGROUND

- In the early 1960s, the General Electric Company published information on Integrated Data Store (IDS), a data base system based on the network model that was later to serve as the foundation for the standards efforts of the CODASYL Data Base Task Group. The IDS document was circulated within IBM among those responsible for planning and development of the programming systems support for IBM's new product line (System/360). The response from one corporate planner was: "We don't need this, we have ISAM (Indexed Sequential Access Method)." This singular lack of understanding of the significance of DBMS within IBM is pointed out to illustrate the environment which gave rise to later DBMS developments.

  - ISAM was proposed as a corporate standard for internal IBM systems development in the 1960s, but performance problems made it impossible to develop major IBM operational or planning systems using that access method. (There are serious questions whether some large IBM information systems, developed in the 1960s using BDAM, can even be converted to a supported data base system such as IMS, much less a relational system.)

  - However, it should not be assumed that the IBM corporate fascination with ISAM precluded other work related to DBMS. Quite the contrary -

GIS, IMS, and CICS were presented to customers as the "planned" solution to the DBMS problems being raised by advocates of IDS in IBM user groups (GUIDE and SHARE). Unfortunately this trio of "solutions" had been developed independently outside the normal IBM software development process, and they did not fit together very well.

- To complicate the situation, GIS, IMS, and CICS were not the only data base-oriented systems being developed internally in IBM - they just happened to get announced for customer use. The IBM on-line, personnel data base was brought up using MIS/360 (an inverted file system similar to RAMIS) and was developed and used internally. However, it was never made available to customers.

- There were so many internal IBM data base efforts (with resulting technical, emotional, and political arguments) that internal technical conferences on DBMSs were being held in the late 1960s.

- It was at this time that E.F. Codd originated the relational model at IBM's San Jose Research Laboratory. The data base wars at IBM might have remained internal if Codd hadn't decided to publish a number of papers through the ACM. (The first was "A Relational Model For Large Shared Data Banks," Communications of the ACM, June 1970.) These days it is seldom that a single individual assumes such total responsibility for a major contribution in the computing industry, but the relational model and Codd have become practically synonymous.

B. DEFINITION AND REVIEW OF RELATIONAL MODEL

- Since Codd is so closely identified with the relational model, it seems only reasonable to accept his definitions of the relational model and what constitutes a relational data base system. He provided definitions when making the 1981 ACM Turing Award Lecture in November 1981.

- The title of the lecture was Relational Data base: A Practical Foundation for Productivity. It presented relational data base management as a foundation for attacking the productivity problem from two approaches:

  - Providing end users with direct access to the information stored in computers.

  - Increasing the productivity of data processing professionals in the development of applications programs.

- It was inevitable that there would be a rush to implement data base systems employing the relational model. However, it was precisely for this reason that Codd drew a clear line between relational and non-relational data base systems by providing precise definitions during his lecture. The relational model is of value because it provides a sound theoretical foundation for data base organization and management; it deserves the attention to definition which it has been given.

- Since all information in a relational data base is represented in tabular form, even the name "relational model" has been questioned. Two reasons have been given for the nomenclature:

  - The term "relational" was specifically selected to counter the then-popular opinion that a relationship between two or more objects had to be represented by a linked structure.

  - Tables were considered to be a lower level of abstraction than relations, since tables imply that positional addressing is inherent and fail to convey that the information content of a table is independent of row order (in other words, tables are not key sorted).

- There were three fundamental objectives which led to the development of the relational model:

  - The first objective was to achieve data independence by defining a clear boundary between the logical and physical aspects of data base management. (Anyone who has ever been exposed to early inverted file systems can appreciate this.)

  - Of equal importance was the provision of a simple structure which could be easily understood by end users, programmers, and data base administrators. Communicability of data base structures is especially important in today's environment where increased emphasis is being placed on information centers, prototyping, and user systems development.

  - The third major objective was to introduce high level language concepts (relational algebra), which would facilitate set-processing and relieve users of the data base from being concerned with the handling of individual records when multiple sets of records are being processed. (Codd emphatically rejects the use of iterative or recursive statements in defining languages that implement the relational algebra.)

- Any data model consists of three fundamental components: 1) a defined set of data structure types; 2) a collection of operators or rules of inference to derive, modify, or retrieve data from the defined data structure types; and 3) a general set of integrity rules that implicitly or explicitly defines a "set of consistent data base states, or changes of state, or both." In the relational model these fundamental components can be defined as follows:

  - The data structure types (and terminology) of the relational model consist of the following:

- Tables contain only one record type.

- Records (rows within a table) have a fixed number of fields (all of which are explicitly named). Records must be unique (no duplicates are allowed) and may come in any order (there is no predetermined sequence).

- Fields are distinct (no repeating groups are allowed).

- Domains represent a range of possible field values (such as employee number). Domains may be used for many different field types and may become the source of values for many different columns in the same or different tables.

- A relation normally refers to a table or record type.

- A tuple is a table row or record occurrence - a group of related fields.

- An attribute is a column name or field type.

- An element is equivalent to a field.

- Degree refers to the number of columns in a table.

- Cardinality refers to the number of rows in a table.

- Binary relations are tables with two columns.

- N-ary relations are tables with N columns.

- An N-tuple is a record from a table with N columns.

- A candidate key uniquely identifies normalized record instances of a given type. Each instance of the record must have a different value on the key, and no attribute in the key can be discarded without destroying the candidate key's ability to locate a single record instance.

- A primary key is used to uniquely identify a record instance or other data grouping.

- At the most fundamental level, the set-processing capability defined for use with relational data base structures is referred to as the relational algebra. The operands are whole relations, any operation results in a new table, and the primary purpose of relational processing is loop-avoidance. The relational algebra is not intended as a standard language, but the basic operators can be used to assess the completeness of implementation of relational systems. The basic operators are:

  - Select takes one relation (table) as an operand and produces a new relation (table) consisting of selected tuples (rows) of the first.

  - Project transforms one relation (table) into a new one consisting of selected attributes (columns) of the first one.

  - Join takes two relations (tables) as operands and produces a third consisting of the rows of the first concatenated with the rows of the second; join is performed only when specified columns of the first and second have matching values. (If redundancy in columns is removed, the operator is referred to as natural join; otherwise it is referred to an equi-join.)

- The general integrity rules in relational data base theory are five principal normal forms which can serve as guidelines for data base design regardless of whether an RDBS is being used. Briefly described, the five normal forms for relational data bases are:

- First normal form specifies that all occurrences of a record type must contain the same number of fields. First normal form excludes repeating fields and groups since relational data base theory does not deal with records that have a variable number of fields.

- Second and third normal forms deal with the relationship between key and non-key fields: a non-key field must provide "a fact about the key, the whole key, and nothing but the key." The record must also conform to first normal form. Second normal form is violated when a non-key field is a fact about a subset of the key. Third normal form is violated when a non-key field is a fact about another non-key field.

- Two additional points must be made about second and third normal forms: 1) both deal only with "single-valued facts" of either a one-to-one or a one-to-many relationship, and 2) both are defined in terms of functional dependencies, which essentially means that a field X is "functionally dependent" on field Y if for this field it is invalid to have two records with the same Y value but different X values.

- Fourth and fifth normal forms deal with "multivalued" facts, which can be defined as many-to-one or many-to-many relationships (as contrasted to the single-valued facts associated with second and third normal forms). Fourth normal form must conform to third normal form and a record type should not contain two or more independent facts about an entity. Fifth normal form essentially covers cases where information can be reconstructed from smaller pieces of information that can be maintained with less redundancy.

- The process of normalization essentially provides for decomposition of records that are in violation of a particular normal form into separate records that do conform to the definitions. The normalization rules (properly applied) prevent many update anomalies and data inconsistencies.

- The support of both entity and referential integrity in the implementation of relational data base systems is both important and challenging.

- The above description of the objectives and components of the relational model are intended as summary information for those who already have some acquaintence with relational theory. A detailed explanation of the relational model is well beyond the scope of this study. Indeed, more than ten years after the publication of Codd's initial paper, relational theory continues to challenge simple descriptions, despite numerous attempts in technical journals. (For example, Communications of the ACM, February 1983, contained a paper titled "A Simple Guide to Five Normal Forms in Relational Data Base Theory" by William Kent of IBM.)

- Clear definition has been further complicated by both experimental and commercial development of relational data base systems, and Codd has addressed the problems of classification (and clarification of terminology) associated with such development efforts.

C. RELATIONAL DEVELOPMENTS AND STATUS

- Since the relational model calls for a particular type of set processing which Codd calls relational processing, as well as relational structures, this capability is considered the key in drawing the line between relational and non-relational systems. Specifically, in order for a data base management system to be considered relational it must support:

- Tables without user-visable navigational links between them, and

- A data sublanguage with at least the relational processing capability of performing the transformations specified by the SELECT, PROJECT, and unrestricted JOIN operators of the relational algebra without resorting to commands for iteration or recursion. (Unrestricted JOIN refers to possible implementation restrictions, such as attributes having to have the same name or a predefined access path.)

- It is suggested that a DBMS that does not support relational processing be considered non-relational, but could be classified as tabular if it supports tables without user-visable navigation links between tables. (This classification is preferred to "semi-relational," which has sometimes been employed.)

- It should be noted that a system may be classified as an RDBS without supporting the following: 1) the rules for entity integrity and referential integrity, and 2) the full relational algebra (three-valued predicate logic with a single kind of rule). Systems which do support these two parts of the model are classified as fully relational.

- The packaging of the relational processing capability is not restricted by the definitions which are given. For example, the INGRES system of Relational Technology, Inc. incorporates all three operators (SELECT, PROJECT, JOIN) in one statement (RETRIEVE of the QUEL language) but yet still qualifies as relational.

- A variety of end-user languages can be developed since the data sublanguages are left open in relation to both extended development and interface to host languages such as COBOL, FORTRAN, APL, etc. In many ways, these user-oriented sublanguages are more important than the underlying data models because they will determine the acceptance of various RDBSs.

- String languages (such as QUEL or SQL) and two-dimensional, screen-oriented languages (such as QUBE) have been developed for specific implementations of RDBSs.

- However, some relational systems (such as System R and INGRES) have implemented a data sublanguage which can be used either embedded within a host language or interactively from a terminal. Codd sees substantial advantages for such "double-mode" languages:

  . Application programmers can separately debug from a terminal the data base statements they wish to embed in their programs.

  . Such a language facilitates communications among programmers, analysts, data base administrators, and end users. (A tremendous advantage - or even necessity - when implementing information-center concepts.)

  . The "frivolous distinctions" between languages is a burden on users who must work in both modes.

- The availability of a double-mode language is considered such a productivity enhancer that a separate classification called uniform relational is recommended for systems that have implemented this feature. Conversely, an RDBS that does not implement a double-mode language is called non-uniform relational.

- Under any circumstances, there are problems associated with exactly how a data sublanguage with relational processing capability can be effectively integrated with a host language that is oriented toward the serial processing of individual records. There are two basic solutions to this problem:

- Derive a relation in the form of a file that can be read record-by-record by the host language; leave the delivery of the records up to the host language file system.

- Or, the data sublanguage may keep control of record delivery and provide record-by-record access to the program written in the host language.

- Codd goes to some length to assure us that the latter does not violate the relational definition: "It is important to note that in advancing a cursor (part of the System R implementation) over a derived relation, the programmer is not engaging in navigation to some target area. The derived relation is itself the target data!"

- Regardless of how the languages interface, there are obvious performance ramifications in communicating between two (or more) languages which expect data to be stored, viewed, accessed, and/or processed in different ways.

- Performance is a critical factor in how RDBSs are developing. Future performance problems which are not fully understood are either being ignored or played down by those who want to speed relational development. The following general observations concerning performance are intended to provide some understanding of the problem:

  - IBM has expended tremendous resources in implementing "research prototype" relational systems. The Peterlee Relational Test Vehicle was developed in the IBM UK Scientific Center and System R was developed in the San Jose Research Center. While performance is always denied as a stated objective of a research prototype, it became an issue with System R. Reports started to leak out during development:

- "I didn't think it was possible, but it (System R) makes IMS performance look great." (A comment concerning early System R performance.)

- "Being better than awful can still be pretty bad." (A comment after a major effort to improve System R performance.)

- Reports such as the above continued over a period of years, and the general conclusion reached was that RDBSs could not achieve acceptable performance when handling "large data bases."

- By 1981, even Codd was forced to conclude that IMS would: "probably be around until the year 2000."

- The standard position of most relational proponents today is that there is "no intrinsic reason relational systems cannot achieve acceptable performance." Perhaps not, but this opinion should not be automatically accepted.

- The JOIN operation conceptually operates as follows:

  - Take the first row from the first table and try to find a row in the second table with a matching value.

  - When a match is found, put the two rows together, forming one new row.

  - Continue until the second table is exhausted.

  - Take the next row from the first table and search the second table for another match.

- Repeat until the first table is also exhausted. The second table has now been searched as many times as there are rows in the first table.

- This description of the operation of JOIN is continued in "A Primer on Relational Data Base Concepts," IBM Systems Journal, Volume 20, Number 1, 1981, G. Sandberg:

  - "The method of operation for a join is very time-consuming and expensive if implemented directly as described. That has been a criticism of relational systems since the beginning. However, improved techniques in areas of query optimization and indexing are developing . . . . Thus, in the join operation previously discussed, if there were an index on a column in the second table, only the index might have to be searched. And for some rows in the first table, no search would be required in the second table at all. Further, if there were also an index on a column in the first table, the search for equal values could be performed entirely in the indexes. The data base system may also keep statistics about actual or intended usage, in order to optimize the search order internally. It now seems that improved optimization methods are sufficiently developed to make possible large-scale relational testing."

  - It is INPUT's opinion that the testing of improved optimization methods on large-scale relational data bases may indeed improve performance, but that does not change the fact that the operation of JOIN is an intrinsic performance problem with the relational model. (As are the data structures themselves.) Being "better than awful" does not make the problems go away.

- In addition, it should be pointed out that RDBSs for large data bases may have to operate with large IBM operating systems that have not been especially effective at implementing "improved optimization methods." Just a few examples will suffice to make the point:

  - There is no intrinsic reason an effective indexed sequential access method cannot be developed, but ISAM hardly qualifies as a good example.

  - Binary searching of tables was a well-known technique for years before IBM implemented OS/MVT with serial searches on frequently used tables within the operating system. (IBM systems have generally been sloppy in table handling and relational systems are tabular.)

  - Keeping statistics on usage can be a good idea but not if it starts to consume more processor resources than those used in problem program execution.

  - IBM has been quoted as stating that an "average transaction" requires 500,000 machine cycles to process. It will be difficult to accept any DBMS that increases this burden.

- The potential optimization techniques mentioned above also point to an intrinsic weakness in the System/360-370 architecture and its 32-bit word (other than the obvious addressing problems which are still being resolved).

  - The limited number of bits precluded instruction formats that might have provided both flexibility and the ordering of indexing within individual instructions.

- A conscious design decision was made not to provide indirect addressing because a bit was not available.

- Both of the above capabilities were known to provide effective (and efficient) means of searching indices and/or keys, but a trade-off was made that continues to be costly in table handling (including sorting).

- These hardware deficiencies in current IBM mainframes may require a hardware assist in order to achieve acceptable performance with large-scale implementation of RDBSs.

- The experience with System R has been well documented publicly, both through IBM sources and through the Communications of the ACM. A great deal of credit is due for the openness with which problems have been identified and discussed in the public forum, and some excellent and imaginative work has been done in addressing some of the performance problems. ("A History and Evaluation of System R", Communications of the ACM, October 1981, is especially recommended for those desiring more detailed information.) However, one is left with the distinct impression that there are still a lot of open questions concerning performance of RDBSs in a large-scale production environment.

- The current implementation status of RDBSs may be summarized briefly as follows:

  - IBM has cautiously extended its SQL/DS RDBS to the VM operating environment with Release 2 in January 1983. This met with the following reactions:

    - "I think we will see it become a standard for IBM DBMS programs, and for other systems as well." (From the president of Oracle Corporation which has a SQL-compatible product.)

- "In the immediate future, we will see the two systems (SQL and IMS) working in a complementary fashion. There are reasons for using one or the other with specific applications. Users could have both at a single location . . ." (an unnamed IBM spokesperson).

- Despite the above, even IBM does not advocate SQL use for large-production data bases.

- Because of both hardware design and less operating systems overhead, relational data base systems such as INGRES (Relational Technology, Inc.) are being implemented on minicomputers which have superior price-performance (compared to IBM mainframes).

- The ease of use inherent in relational systems is encouraging the development of a wide variety of relational systems for microprocessors (personal computers). Performance problems are solved by inherent limitations on data base size and by design for single use.

- Data base computers (DBCs) that facilitate the set-processing and associative memory capabilities necessary for effective RDBS implementation are beginning to appear in the marketplace. (Britton-Lee and INTEL are the primary examples.)

  - When used in conjunction with mainframes as backend processors, DBCs relieve performance problems by off-loading the RDBS processing and storage management functions, and also provide a clean interface to mainframe systems software.

  - As standalone data base engines, imaginative systems can be developed in conjunction with personal computers. Data base computers can also be connected for distributing data bases.

- Experience with data base computers has been limited, and the development of new systems concepts is slow in an environment where users have adopted a healthy skepticism about new solutions to long-standing problems.

D. USER REACTIONS

- For a number of years, INPUT has asked users questions about their understanding and opinions of hierarchical, network, and relational data models. Essentially, these questions have elicited the fact that few users have a detailed technical knowledge of the models, and only a few have had any strong opinion about them. In fact, until recently there was, in addition to embarrassment about not being informed on such a "hot topic," a general lack of interest in the subject. This was epitomized by one IS manager who stated: "I don't even like to talk about data models - the whole subject bores me."

- This seeming insensitivity to data models should not be interpreted as a lack of interest in data base management systems. IS managers have been concerned about DBMSs for years, but they have been interested primarily in the promise of data base systems and not in the technical aspect of data structuring. This was especially true of the relational model, which until recently was considered an academic exercise. Users were encouraged in this feeling by IBM, which was busy selling IMS against both internal and external competition.

- For those users who took the plunge with IMS, the question of conversion costs from batch to data base was quite sensitive. One executive responded with unusual candor when he stated: "I don't know (how much conversion cost), and I don't think we want to know."

- Recently users have exhibited more interest in data models - if not more detailed technical knowledge. Earlier this year a sample of IS managers was asked for their best perception of hierarchical, network, relational, and other models. The response can be summarized as follows:

  - Network was considered to be designed with the DB administrator in mind, was felt to optimize data base structure and operation, and was identified as being "CODASYL-approved."

  - Hierarchical was felt to facilitate application implementation from a programmer's point of view. This model was classified as a special case of the network model; it was considered less flexible than the relational and less efficient than the network.

  - Relational was felt to be user friendly, flexible, and expensive in terms of hardware system utilization.

  - Users were also asked about a "coexistence" model, and they stated it would accomodate the best of the above. When asked about a set theoretic model (a model accommodating multiple views of data structures), respondents had no knowledge or opinion.

- User reactions seemed to reflect the general marketing and sales promotion of competing DBMSs, rather than the detailed knowledge that would come from serious evaluation of alternatives. In the next two years, users will be confronted with decisions that will require significantly more knowledge of data base structures than can be gained from product announcements and general comments in the trade press.

IV THE PROJECTED ROLE OF THE RELATIONAL MODEL

A. THE PROBLEMS WITH PAST "SOLUTIONS"

- Data base sytems were conceived as a solution to the problems of providing timely, accurate planning and control information from the mass of unorganized data that was becoming available from "computer files." The history of DBMSs has been accompanied by evolving terminology (management information systems, decision support systems, information engineering, information centers, information resource management, etc.), that has essentially repackaged the fundamental promise of timely and accurate information so that it will appear new. The actual experience with DBMSs has generally been fraught with frustration for end users because ready access to quality information always seems to be promised but is seldom provided by the latest breakthrough in terminology.

- A brief review and analysis of the problems associated with past solutions is necessary before today's answers can be evaluated.

  - Early DBMS software was developed and installed before it was realized that the data were either inadequate or not available.

  - It was then determined that the development and management of a data base required a lot of detailed, thankless work which was not especially interesting to the systems and programming people responsible for the software development. It became necessary to invent the data base administrative function.

- Next it was discovered that end users did not really know what they wanted (or needed) in a data base or even understand the "information" they were already using. It became necessary to make a substantial corporate commitment to define the data base(s) required to convert from batch data processing to on-line information systems. The commitment of resources was justified because it was viewed as a one-time investment.

- At the end of a substantial effort to build the central data base came the horrible realization that the requirements had actually changed during implementation (or that even a relatively simple organizational change required great effort to restructure the data base).

- During this time, systems personnel became increasingly involved in implementing the latest systems "solutions" to the problem of providing timely information to the end users. This involvement in seeking the ultimate solution meant that information systems personnel could not respond to current requests for information, and as a result the backlog grew. This caused everyone to focus on the productivity problems in the systems development process.

• It is important to recognize that the investment in developing corporate data bases (and associated software) has contributed substantially to the problems of both data base access and the productivity of data processing professionals. These are the problems Codd wants to solve with the relational model. In other words, the RDBS is intended to solve problems created by its predecessors.

• However, before defining the role of RDBSs in the continuing search for timely management information, specific words of caution are necessary on two currently popular solutions to the problem.

- The first is best defined as the last gasp of the big bang theory of data base development. It re-emphasizes that information (as opposed to data) is a corporate asset and is worth one last, all-out effort to define corporate information requirements. (IBM's BSP, and Information Engineering as defined by Finkelstein and Martin, are examples of the big bang approach). If past experience is any guide, the investment of enormous resources over an extended period of time and the capitalization of the resulting information base (as recommended by some advocates of the big bang theory) should be frightening to even the least prudent IS manager (to say nothing of the corporate controller).

- At the other end of the spectrum, frustrated users are installing personal computers, keeping their personal data bases on floppy disks, and using electronic spreadsheets to generate their own planning and control information. The cost justification and quality of such systems is also highly suspect, even though the resources being expended frequently do not appear in the IS budget. Either the information needed for corporate planning and control is much simpler than we have been led to believe, or a lot of companies are exposing themselves to problems which will make the old-batch "computer files" look like a model of data organization and integrity.

- It appears that the extreme solution of one big data base and infinite personal data bases may be creating tomorrow's problems, and those problems will <u>not</u> be simpler than today's.

    - Flexibility to accommodate change will still be required regardless of how carefully corporate information requirements are analyzed. The hierarchical and network data models do not provide sufficient flexibility, and the relational model has not demonstrated acceptable performance when used for large production data bases.

- The integrity problems associated with the synchronization of distributed data bases have not been solved. Personal data bases will lead to "information conflict" within the organization; central data bases will also be subject to contamination.

- The communications gap between advocates of top-down data base development and those involved with bottom-up development is widening, and philosophical solutions are going to be difficult to sell - much less implement.

- The relational model was specifically designed with the objective of data independence, communicability, and set-processing capabilities. All three of these objectives significantly enhance ease of use by all kinds of users, and facilitates ease of communications among all levels of data base designers. Bridging the communcations gap between corporate data base designers and end users could be the most important contributions of an RDBS, provided the technology is properly applied and not viewed as a universal solution.

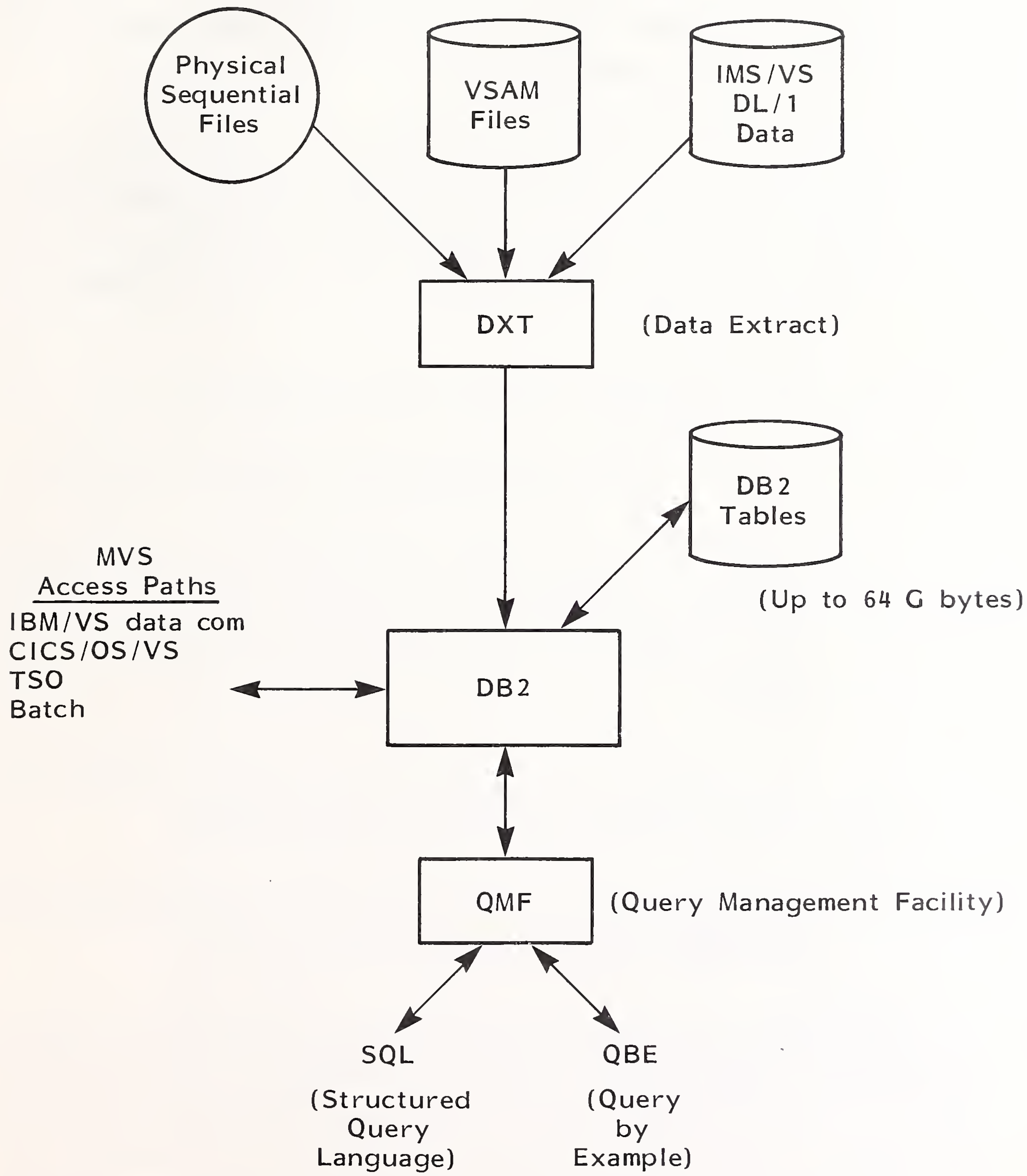
B. THE PLANNED ROLE OF RDBSs

- In June, 1983, IBM announced Database 2 (DB2), which is a relational data base management system for MVS/XA and MVS/370 architectures. This announcement and related support programs give some indication of IBM's planned role for the relational model, but care should be exercized in assuming that DB2 itself is the final solution to many of the historic problems outlined in the previous section of this report.

- DB2 was developed to support large data bases of 64 gigabytes. The announced architectural environment is depicted in Exhibit IV-1.

EXHIBIT IV-1

DB2 GENERAL ARCHITECTURE



- MVS users may access DB2 through the IMS/VS data communications feature, CICS/OS/VS, TSO, and in-batch mode.

- The Query Management Facility (QMF) allows users to extract, manipulate and generate interactively reports using IBM's Structured Query Language (SQL) and Query-by-Example (QBE). Data definition functions are performed using SQL.

- Data Extract (DXT) "extracts selected operational data" from IMS/VS or DL1 data bases, VSAM data sets, and sequential (SAM) files and prepares them for loading into DB2. DXT is "designed for programmers or users" to facilitate extract requests that are supported as follows:

  . One or two views of the data when it's extracted.

  . An OS/VS DB/DC data dictionary can be used for stored data.

  . Dialogs under QMF allow interactive request construction and submission, and consist of Interactive System Productivities Facility panels that guide the process of creating an extract request.

  . JCL prompts user-configurable model extract statements and request submission capabilities are also included.

  . In general, the extraction of data from existing files and the building of DB2 data bases are made quite easy.

- Generally speaking, the IBM DB2 announcement was greeted as a solid endorsement of RDBSs and as a triumph for both ease of use and end users in general. Some observers went so far as to state that it would permit user data bases "to slide into the relational world a step at a time" and one predicted IMS data structures would disappear within five or ten years. However, one industry pundit was disappointed because it was not possible to make queries directly against a production data base and felt that other systems such as Cullinet's IDMS/R were more determined to demolish the distinction between production and end-user data bases.

- INPUT feels that the distinction between production and end-user data bases is still warranted because many of the problems of distributed data bases have not yet been solved. Specifically, the normal forms of the relational model should be used as guidelines for record design regardless of either the level of data base being considered (production or end user) or the physical data structure. However, this is not generally understood and any conversion is going to take time.

- It is INPUT's feeling that even the loose coupling of production and end-user data bases represented by DB2 will encourage updating up and down the hierarchy regardless of the original intent. This presents IS management and data base administrators with a dual challenge, but it may force progress.

    - At the present time, IS managers generally feel that they will be forced to provide authorized end users with data from central files. With the extract facilities associated with DB2, the tools to force distribution have been put into the end users' hands.

    - Most IS managers (or DB administrators) are currently insisting that they will exercise their responsibility for data integrity by controlling updates to the central (production) data base. End users are going to insist on this capability, even though there is currently no provision for easy update of production files under DB2 (a strength, not a weakness). It is probable that as the relational model is better understood, the advantages of relational normalization will become apparent to those responsible for the production data bases.

- While everyone argues about production versus end-user data bases, INPUT's research (prior to the DB2 announcement) uncovered some profound observations concerning the use of the relational model against archival files that currently contain a vast reservoir of valuable data that is stored sequentially on magnetic tape and is generally inaccessible for most practical purposes. The following paraphrased comments were received from the respondent:

  - "If archival files are defined relationally they are structured for unanticipated use, and can be used effectively with the relational algebra and/or calculus to project appropriate planning files."

  - "Depending upon the particular circumstances, it may be worthwhile to breathe life back into archival files by redefining them relationally. It is like extracting gold from a dump - there comes a time when it pays off."

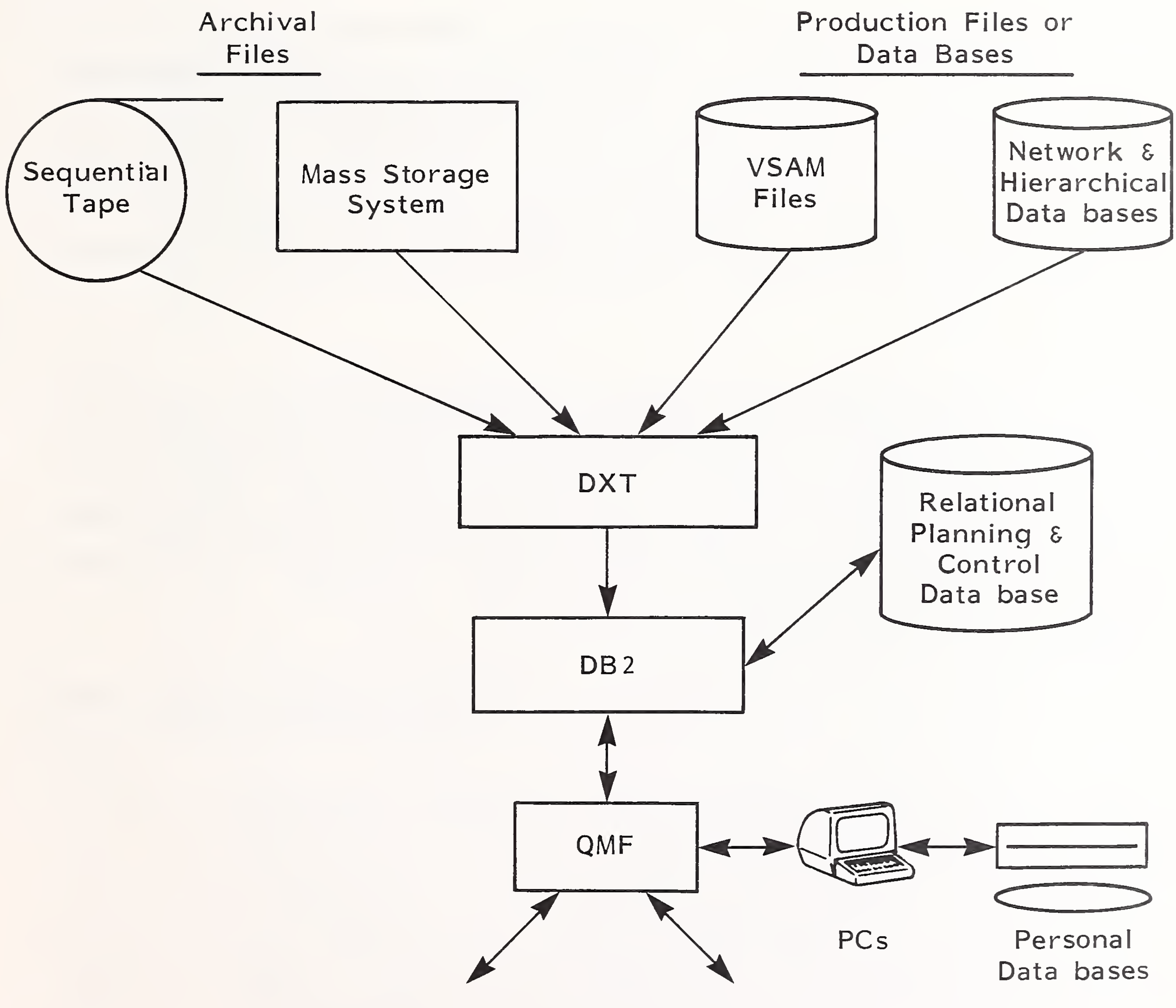
- It is INPUT's opinion that these are important observations, especially when considering the possibility of future archiving onto optical disk (see <u>Impact of Upcoming Optical Memory Systems</u>, April 1983). With the announcement of DB2 and DXT, the tools are now available to provide relatively economic access to data that otherwise would be lost because it cannot be readily (and flexibly) accessed. However, the "extraction of gold from a dump" can be prohibitively expensive and this brings us to the question of potential DB2 use and performance.

C. PRACTICAL CONSTRAINTS ON RDBS USE

- The fact that it has been made relatively painless for end users (or IS professionals, for that matter) to tap into production or archival data bases using relational systems (such as DB2) may solve many past problems attributable to DBMSs, but it will not be an unmixed blessing. It is not difficult to envision an overall DBMS operating environment that could prove quite expensive in terms of on-line storage costs, as shown in Exhibit IV-2.

EXHIBIT IV-2

POTENTIAL DB2 OPERATING ENVIRONMENT



- By making it easy to extract data from archival and productive data bases and files, it will be possible to build "large shared-data banks" for purposes of planning and control. These banks will employ the relational model just as Codd envisioned in his 1970 paper. ("A Relational Model Of Data For Large Shared Data Bank." Communications of the ACM, June 1970.)

- Even with reasonable control (by the data base administrator), this will result in duplication of substantial portions of current production data bases and archival files. Assuming that the operating environment is easy to use (and that the users are willing to foot the bill), then there will probably be substantial overlap across the planning data bases as using departments and/or individuals select, joint and project their own working copies of the planning files.

- In addition, another level of duplication will occur as individuals extract personnel files (data bases) from higher levels for use on their own personal computers. While personal files on floppy disks may be of negligible concern, investment in hard disk for PCs can be substantial, and the time is coming when PCs may have copies of entire planning data bases available on optical disk.

- For years IBM discouraged the development of distributed processing by pointing out the evils of duplicate data bases. Emphasis was placed on getting everything under firm control on the large host computers where common data could be shared by all. DB2 makes it inevitable that there will be a tremendous amount of duplicated data throughout the overall data base environment (which will sell an awful lot of disk storage). There will be duplicated data because most data processing organizations have failed to control on-line storage in far-simpler operating environments.

- It is true that storage costs are dropping rapidly. However, the tools are now available to assure that demands for on-line storage will increase rapidly enough to more than compensate for the advances in storage technology (in terms of lower cost-per-bit). In addition, there are processing costs associated with the creation and maintenance of all of those overlapping data bases, and in the use of the relational model.

- It has been stated that IMS succeeded in "burning CPU cycles beyond IBM's wildest dreams," and it probably can be established that increased use of DBMSs (and especially IMS) has driven the demand for ever larger main-frames. Now consider the fact that RDBSs (specifically System R) have had performance which makes IMS look good, and it becomes possible to understand why even IBM has been reluctant to announce an RDBS for large data bases.

- While it is obviously impossible to evaluate DB2 performance at this time, it is worthwhile to look more closely at the experience with System R since DB2 is based on that experimental effort. As mentioned previously, IBM has been rather open on publishing the results of the System R effort, and it is possible to gain considerable insight into the performance problems which were encountered, as well as the implementation approaches which were taken during System R development. The following is a brief summary of critical problems and approaches:

  - An important "discovery" in the development of the System R prototype was that the cost (overhead and performance) of the relational access method was best measured by the number of I/Os rather than the number of tuples fetched. (Why this was a great revelation in the mid-1970s has not been explained, but it is assumed to have something to do with the backgrounds of those engaged in experimental projects.) Among the observations and conclusions that were reached are the following:

- Manipulation of "tuple identifiers" (TIDs) could be extremely expensive - especially when they were on direct-access storage (therefore requiring additional I/O).

- It was important to cluster related tuples together on a physical page "so that several related tuples could be fetched with a single I/O."

- "Strong domains separately from tuples causes (sic) many extra I/Os to be done in retrieving data values."

- However, the fascination with I/O activity as a measure did not conceal the fact that the prototype implementation was CPU-bound on a typical query (the prototype was a one-user at a time system implemented on an IBM 370/168). It was concluded that the optimizer should consider CPU time as well as I/O activity. This resulted in the design of a Relational Data System (RDS) that compiled very high-level SQL statements into relatively "compact, efficient routines in machine language." These routines became the access modules that were called from the application program.

- These modules became part of the Research Storage System (RSS), which is the relational access method eventually employed with the full-function multiuser version of System R. This system was thoroughly tested at a number of IBM installations. RSS access paths had the following general characteristics:

  - Based on experience with the protype system, where separate domains were maintained for data values, RSS chose to store data values in the individual records of the data base. This choice resulted in variable length records which were generally longer than comparable records in the prototype system.

- The access paths provided by RSS permitted index scans (using B-trees), link scans (which traverse from one record to another) and relation scans (which scan the tables as they are laid out in physical storage).

- Search arguments could be employed to limit the records returned, and a built-in sort was also made available.

- The prototype system also identified the JOIN formulation as being of critical importance, and the full-function systems took advantage of independent research that studied ten methods of joining together tables, based on the use of indexes, sorting, physical pointers, and TIDs. Two were selected for implementation as being optimal in most circumstances:

  - Join Method 1 scans over the qualifying rows of table A and for each row fetches the matching row of table B (most usually employing an index on table B).

  - Join Method 2 sorts the qualifying rows of tables A and B in order of their join field, and then scans over the sorted lists and merges them by matching values. (This method is frequently employed when no suitable index exists.)

- In addition, System R also contained well thought out subsystems for authorization, recovery, and locking (to prevent interference among concurrent users). Generally speaking, the full-function system implementation demonstrated careful analysis and sensitivity to performance requirements.

- The evaluations of System R at internal IBM installations resulted in the following observations concerning performance. (The published sources for this information are contained in the publications cited in the Executive Summary.)

- "In general, the experimental data bases used with System R were smaller than one 3330-disk pack (200 megabytes) and were typically accessed by fewer than 10 concurrent users. As might be expected, interactive response slowed down during the execution of very complex SQL statements involving joins of several tables. This performance degradation must be traded off against the advantages of normalization in which large data base tables are broken into smaller parts to avoid redundancy, and then joined back together by the view mechanism or user applications."

- The generation of machine language routines to execute SQL statements was found to be especially beneficial for short, repetitive transactions but less clear in an ad hoc query environment. Simple analysis revealed that approximately 20% additional CPU time was required for code generation. This was on top of the time required for passing and access path selection (which are required by both interpretive and compiler-oriented query processors). It was concluded that this increase would not be perceptible to an end user, but this type of reasoning can lead to sloppy implementation, which can have an effect on response time in highly interactive environments with many concurrent users.

- Concerning access paths, it was concluded that B-tree indexes could "be used very efficiently in queries and transactions which access many records, but hashing and links (which were implemented only internal to the system and not for applications data bases) would have enhanced the performance of 'canned transactions' that access only a few records." (This conclusion is important because the savings is in I/O activities to retrieve B-tree pages.)

- The optimizer developed for System R was considered to be good enough to at least order the cost of the various access paths selected for a variety of SQL statements. However, it was felt that more experience was required with access path selection because good access paths might be overlooked when they were not included in the optimizer's repertoire.

- The recovery subsystem was found to have a perceptible impact on performance because a "shadow page" was required to be maintained for each updated page. The reasons for the impact are significant:

  - Each updated page is written out to a new location on the disk, and this limits the ability to cluster related pages and minimizes disk arm movement for sequential applications. (In other words the recovery system had a negative effect on the benefits of a performance-improvement objective.)

  - The "old" and "new" versions must be maintained in a directory, and it was stated that for large data bases the directory could become large enough to require a paging mechanism of its own. (XA may help in this regard, but the problem is apparent.)

  - The periodic checkpoints required for "old" and "new" page pointers generate I/O activity and consume CPU time.

  - A possible alternative solution for future implementation was mentioned. It would substitute a simple log of all data base updates. (The need for minimizing I/O activity when writing out the log was anticipated in the evaluation - suggesting the cure might be worse than the disease.)

- The locking subsystem had varying performance results based on the level of isolation selected, but a serious performance problem that became known as the "convoy phenomenon" was isolated. This phenomenon is important because it demonstrates the unanticipated problems that always develop when interfacing major new software systems with current operating systems.

- Essentially, high-traffic locks in System R tended to be serialized by the operating system dispatcher, and under certain circumstances a process would go into a long "time-slice wait" while holding a high-priority lock. Then other processes would become enqueued behind the "sleeping process."

- The net result, under certain circumstances, was to have requests for locks arriving in less time than the dispatching overhead. That resulted in "thrashing" (our old friend of early VS days) where the operating system starts to use virtually all the CPU cycles.

- Although a "solution" was developed for the problem, it was highly qualified with terms like "it is highly probable that process P will not be holding the lock" and "if a convoy should ever form, it will most likely evaporate."

- There are extremely complex interrelationships between a DBMS (especially one as flexible and interactive as a relational system) and the operations system under which it operates. It is highly probable that other "phenomena" will occur when RDBSs are installed in complex user environments.

- It is INPUT's opinion that there will be a tendency for production data bases to "slide away from IMS" when DB2 becomes available, but it is probable that expense (in terms of both storage and larger mainframes) and performance (responsiveness and systems degradation) will constrain the anticipated rush by large users to relational systems.

- There is a big difference between experimental data bases of less than 200 megabytes and 10 concurrent users, and DB2's potential 64 gigabytes and hundreds of concurrent users.

- Even if the performance problems discovered on System R have been carefully analyzed and imaginative solutions implemented in DB2 (which can by no means be automatically assumed), it is felt that installations of large relational data bases in a variety of end-user environments will uncover performance problems not anticipated by the developers.

- It is also doubtful that the original extract programs (DXT) will be effectively implemented. There is even some indication that "relational purists" would prefer to forget sequential files exist. For example, an extraction from a sequential tape file would be assumed by DB2 to be like any unordered relational table. (While it is hoped that such technical nitpicking does not spill over into a released product, stranger things have happened in the past.)

- In addition, we must once again point out that the hardware architecture of IBM mainframe systems does not provide the richness and flexibility of memory addressing and protection which would easily facilitate improved performance.

- Assuming RDBSs do address data base access and productivity problems in an effective manner, and that acceptance (or applicability) may be slowed by performance problems associated with complex software interactions, it is probable that hardware implementations of RDBSs will become increasingly attractive.

D. HARDWARE IMPLEMENTATION

- The concept of data base machines (DBMs) has been around for a number of years and is based on the attractiveness of separating data from programs and unloading general purpose mainframes from the increasing processing burden of data base systems. There are fundamental reasons DBMs should be more cost effective than general purpose mainframes in handling data management functions:

    - The DBM can be specialized to perform the relatively simple tasks of compare, sort, and merge, which are the essense of much data base processing.

    - The Von Neumann architecture of most current computers need not be a restriction in the design of DBMs.

    - Associative memories, which have been researched for nearly 20 years, can be especially effective when combined with DBMs. The associative memory can be loaded with blocks of the data base and swept simultaneously in search of key values. (Corem's Synfobase is reputed to be the first commercially available DBM that employs associative memory, but a number of universities and vendors have developed prototype systems.)

    - The relational model is especially well suited for implementation on data base machines because both emphasize data independence as a primary objective. All six data base machines that are currently commercially available employ the relational model.

    - By establishing clear data independence in both software and hardware implementation of the relational model, it should be possible to avoid the complex software interfaces with host operating systems that were identified as major performance problems.

- Without going into a detailed description of currently available DBMs, there are three general systems configurations:

  - To effect the improved performance mentioned above, the backend data base machine, as shown in Exhibit IV-3, serves to off-load the DBMS functions from the host. (The backend data base machine depicted is based on the relational model, but some, such as the Intel Data Base Processor (IDBP), support the network and hierarchical models as well). The DBM itself may look like a minicomputer with separate microprocessors to perform specific functions.

    - All Britton-Lee models include a dedicated microprocessor to perform the DBMS functions that normally would be handled in software. All models also offer a special-purpose micro (data base accelerator) with a three-stage pipeline architecture that can scan and filter out relevant data between the time the disk operation is initiated and the time it is transferred to the data base processor.

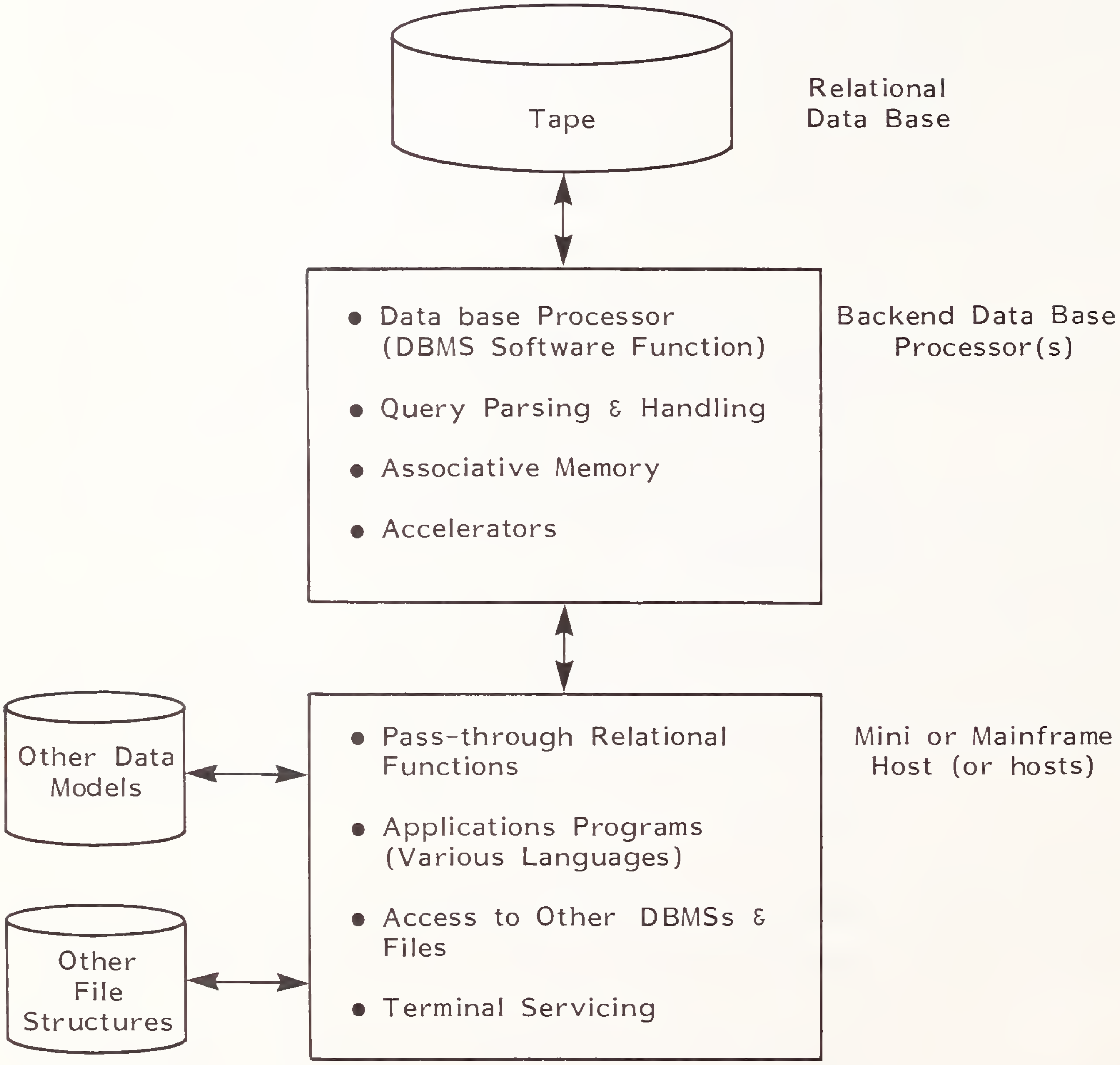
    - Query parsing and handling are currently implemented primarily in software for purposes of OEM sales and tailoring to various host systems (software and hardware). However, a dedicated micro could be used to good effect as query languages became more stable.

    - Addition of associative memories to backend processors will become increasingly attractive with advances in LSI technology.

    - The architectural distribution of processing represented by backend data base machines is also inherent in the DBMs themselves.

EXHIBIT IV-3

BACKEND DATA BASE MACHINE
(Relational)



- At the end of the System R evaluation, IBM set "two major foci" for continuing research: adaption of System R to a distributed data base environment, and extension of the optimizer algorithms to include a broader set of access paths. Data base machines not only serve to supplement host performance, but they are especially well suited to the distributed data base environment, as shown in Exhibit IV-4.

  . Data base machines at communications network nodes would perform all of the functions of a backend DBM, plus local transaction processing and communications with the host or other distributed DBMs.

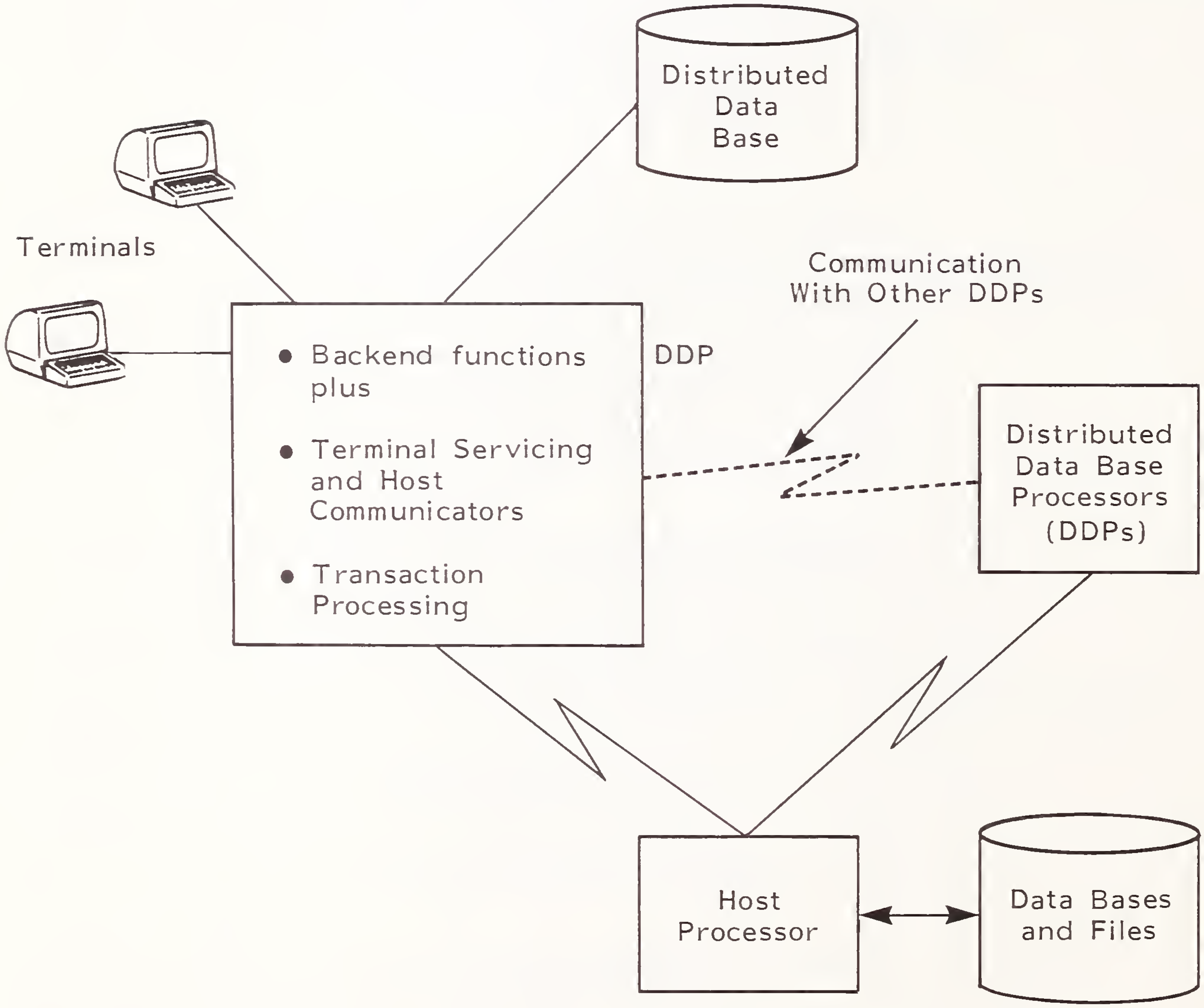
  . The Mega/Net Distributed Data Base Machine which is being marketed primarily in Europe claims compatability with X.25 and Ethernet. It can also serve as a backend data base machine, and as a standalone transaction processor. (Theoretically it can support up to 500 G bytes of disk storage, 850 I/O ports, and up to 128 communications lines.)

- INPUT has projected both the architectural and geographic distribution of processing using DBMs for a number of years. It would appear that the productivity advantages of RDBSs (plus the performance improvements which will be required to implement such systems under MVS) will give impetus to the development of such systems. The standalone transaction processor utilizing the RDBS is practically a fall-out of the other two, but the advantages of the relational model on PCs would have forced such development under any circumstances.

• It is INPUT's opinion that IBM's announcement of DB2 will result in the architectural distribution of processing in IBM hardware systems shortly after the release of DB2 in 1984. It is immaterial whether this distribution takes the form of a backend DBM, an intelligent controller, or is hidden under the covers of a new generation of mainframes - the advantages, and even neces-

EXHIBIT IV-4

DISTRIBUTED DATA BASE PROCESSOR (DDP)



sity, of hardware implementation of DBMSs will become apparent. The general architecture of the hardware versions of DB2 is depicted in Exhibit IV-5.

- The DBM could be a "multi-micro" architecture similar to some of the backend DBMs described above, but it is more likely to take the form of an intelligent controller.

- DXT is an especially good candidate for a specialized micro-based processor within the DBM.

  . The possibility of scanning extracted data values "on the fly" (similar to the Britton-Lee accelerator) would be very appealing.

  . The update capability indicated may not be implemented (in fact, it probably won't be in the original version). In this case, the IMS, VSAM and sequential files would also be connected to the host.

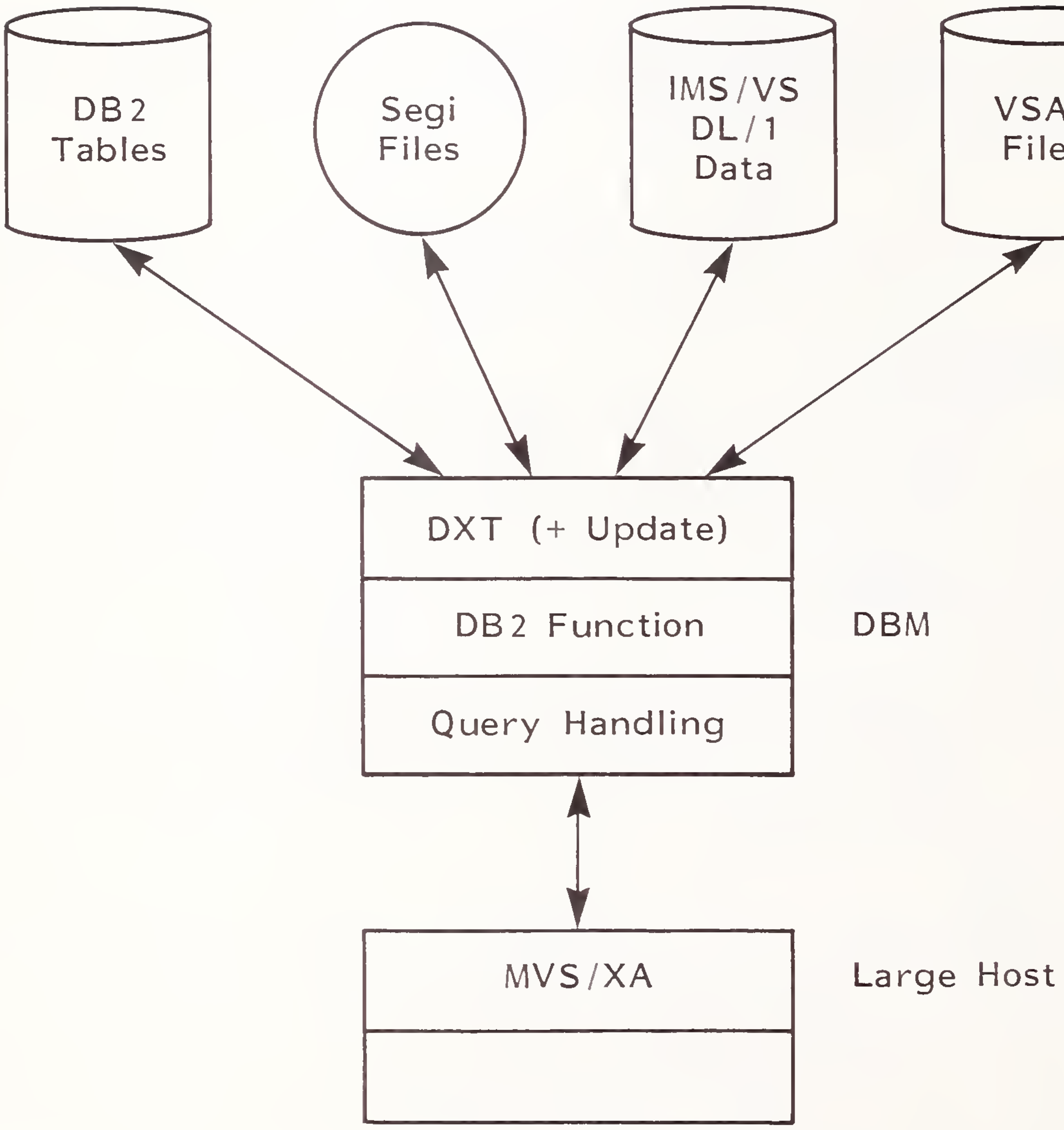
  . If implemented, the update capability would probably be I/O-handling only (controller functions), with the host retaining the access methods.

- However, there is the possibility that a general purpose DBM that supports various data models (including network) could be announced by IBM as a backend. This will depend upon the success of competitive vendors and the ability of IBM to redirect and coordinate competing internal development efforts.

• Under any circumstances, there will be excitement in backend DBMs in the mid-1980s. However, it is doubtful that IBM will encourage the geographic distribution of processing on DBMs. Mainframe host processors for distributed data bases are going to be with us for some time.

EXHIBIT IV-5

HARDWARE IMPLEMENTATION OF DB2



E. NEW INFORMATION MANAGEMENT REQUIREMENTS

- Regardless of the tendency to feel that DBMSs are the ultimate solution to all data processing problems, there is the fact that as data processing is distributed from the top down, office automation is growing from the bottom up. The need to manage all information - data, text, images and voice - has become increasingly important. Today's data base systems (including relational) address only part of the problem.

- Work is just beginning on new techniques which will attempt to bridge the gap between data and information. Surrogate data base models based on the extraction of key words from text have been experimented with, and it is possible that this can be extended to voice messages. Essentially, the surrogate model lends itself to rapid association of words or numbers (data values) with documents or messages (information).

- Those involved with artificial intelligence have identified knowledge acquisition as the critical bottleneck. It is suggested that once the knowledge acquisition problem is solved, it will be found that data and information management problems will remain as the critical factors in the implementation of "knowledge-based" systems. (In this regard, data and information management will determine the success of Japanese fifth-generation computer systems.)

- It is intuitively felt that the relational model and associated hardware developments hold the most promise for bringing these information management problems into focus and closer to solution. By this we mean that RDBS development is tending to inspire hardware/software systems architecture that is more likely to satisfy tomorrow's information system requirements.

F. GUIDELINES FOR MANAGEMENT PLANNING

- It is important that the relational model be understood. The theoretical mathematical orientation and terminology have been barriers to understanding and acceptance. However, uncomplicated descriptions, which will put the model in proper perspective, are becoming available. It is strongly recommended that Codd's definitions be applied in evaluating various implementations of relational systems.

- Assume that a multiple data structure environment will be both necessary and desirable in the foreseeable future. IMS and sequential files will be around for a long time and any plan to install a relational system should take this into consideration.

- The normal forms of the relational model should be used as guides for data base design in the multi-model environment. They can be applied regardless of how the data base is navigated, and they will ease the problems of integrity among data bases (as well as general movement of data among models).

- The inherent performance ramifications of the relational model should be understood by all those involved (systems analysts, programmers, data base administrators, and end users). It is extremely important to consider performance impacts of data base size, level of normalization, and use of the relational algebra and calculus. To install a relational system (such as DB2) in an MVS/XA environment without sensitivity to performance (and appropriate controls on applications and use) will be to court disaster.

    - Study the published performance evaluation of System R in detail.

    - Require an explanation of implementation details of any relational system being considered.

- Devise some simple benchmarks and ask that they be run (JOINS on various size tables would be appropriate).

- Ask for information on instructions executed and I/O activity for various transactions (and functions).

- Establish performance guidelines for RDBS applications and installation. Educate all users in their use.

- Control and evaluate the use of RDBSs from the point of view of performance impact.

- Plan for and control not only central-site installation of RDBSs, but also standalone relational systems for minicomputers and PCs. If ignoring performance of RDBS could prove disastrous, permitting uncontrolled installation of non-compatible "relational systems" could result in catastrophy - there is a difference.

  - Relational systems facilitate accessing data and implementing applications systems, and these capabilities are fundamental in the current emphasis on information centers, decision support systems, and prototyping.

  - However, <u>ad hoc</u> queries and reporting and prototype systems eventually become embedded in the production workload, at which time the responsibility for either installation or conversion is passed to the central IS function.

  - The transfer of these sytems to a central RDBS and associated large data base may prove not only difficult but impossible.

- Consider the potential of RDBSs and DBMs for conventional DP applications, office systems (distributed processing), and even knowledge-based systems, but <u>do not assume they are the final solution to any of the problems</u>.

APPENDIX: RELATED INPUT REPORTS

- Impact of Upcoming Optical Memory Systems, April 1983.
- Personal Computers in the IS Strategy, December 1982.
- Organizing the Information Center, July 1983.
- Personal Computer Software Support, July 1983.

MANAGEMENT PROGRAMS: Designed for clients with a continuing need for information on a range of subjects in a given area.

- Management Planning Program in Information Systems - Provides managers of large computer/communications facilities with timely and accurate information on developments that affect today's decisions and plans for the future.
- Management Planning Program for the Information Services Industry - Provides market forecasts and business information to information services companies to support planning and product decisions.
- Company Analysis and Monitoring Program for the Information Services Industry - Provides immediate access to detailed information on over 3,000 companies offering software, processing services, integrated systems, and professional services in the U.S. and Canada.
- Management Planning Program in Field Service - Provides senior field service managers in the U.S. and Europe with information and data to support their planning and operational decisions.
- On-Target Marketing - A practical, "how-to" methodology for more effective marketing problem solving and planning. Delivered to clients through workshops and/or consulting services.

MULTICLIENT STUDIES: Research shared by a group of sponsors on topics for which there is a need for in-depth, "one-time" information and analysis. A multiclient study typically has a budget of over $200,000, yet the cost to an individual client is usually less than $30,000. Recent studies specified by clients include:

- Selling Personal Computers to Large Corporations
- Improving the Productivity of Systems and Software Implementation
- User Communication Networks and Needs
- Financial Planning Systems Markets: The Next Five Years

CUSTOM STUDIES: Custom studies are sponsored by a single client on a proprietary basis and are used to answer specific questions or to address unique problems. Fees are based on the extent of the research work. Examples of recent assignments include:

- Organizing for Effective Software Development
- Corporate Plan for Utilizing CAD/CAM
- Annual ADAPSO Survey of the Computer Services Industry
- Analysis of Business Services for a Major Financial Institution
- Study of the Specialty Terminal Market
- Study of Disaster Recovery Services
- Analysis of Software Maintenance Issues
- Review of Software Product Market Opportunities
- Analysis of Network User Requirements

About INPUT

INPUT provides planning information, analysis, and recommendations to managers and executives in the information processing industries. Through market research, technology forecasting, and competitive analysis, INPUT supports client management in making informed decisions. Continuing services are provided to users and vendors of computers, communications, and office products and services.

The company carries out continuous and in depth research. Working closely with clients on important issues, INPUT's staff members analyze and interpret the research data, then develop recommendations and innovative ideas to meet clients' needs. Clients receive reports, presentations, access to data on which analyses are based, and continuous consulting.

Many of INPUT's professional staff members have nearly 20 years' experience in their areas of specialization. Most have held senior management positions in operations, marketing, or planning. This expertise enables INPUT to supply practical solutions to complex business problems.

Formed in 1974, INPUT has become a leading international planning services firm. Clients include over 100 of the world's largest and most technically advanced companies.

OFFICES

Headquarters
1943 Landings Drive
Mountain View, CA 94043
(415) 960 3990
Telex 171407

Detroit
220 E. Huron
Suite 209
Ann Arbor, MI 48104
(313) 971 0667

New York
Park 80 Plaza West-1
Saddle Brook, NJ 07662
(201) 368 9471
Telex 134630

United Kingdom
INPUT, Ltd
Airwork House
35 Piccadilly
London, W1V 9PB
England
01-439-8985
Telex 23116

AFFILIATES

France
La Nacelle
2, rue Campagne Premiere
75014 Paris
322 56 46
Telex 220064 X5533

Italy
PGP Sistema SRL
20127 Milano
Via Soperga 36
Italy
Milan 284 2850
Telex 310352

Japan
Overseas Data Service Company, Ltd.
Shugetsu Building
No 12 - 7 Kita Aoyama
3-Chome Minato-ku
Tokyo, 107
Japan
(03) 400 7090
Telex 26487

Sweden
Athena Konsult
P.O. Persson & Co AB
Box 22114
S 104 22 Stockholm
Sweden
08-52 07 20
Telex 17041

INPUT Planning Services for Management

INPUT

MANAGEMENT PLANNING PROGRAM IN INFORMATION SYSTEMS

VENDOR WATCH REPORT

LOCAL AREA NETWORKS: DIRECTIONS AND OPPORTUNITIES

DECEMBER 1983

MANAGEMENT PLANNING PROGRAM IN INFORMATION SYSTEMS

OBJECTIVE: To provide managers of large computer and communications systems with timely and accurate information on developments that affect today's decisions and plans for the future.

DESCRIPTION: Clients of this program receive the following services each year:

- Impact/Planning Support Studies - In-depth reports dealing with the impact on users of projected managerial, personnel, and technological developments over the next five years. Studies include analyses and recommendations.
- Technology and Management Issue Briefs - Analyses of the probable moves of major computer/communications vendors in operating systems, data base/data communications, s ... lue-added networks, personal computer ...
- Residual Value Fo ... dual values of major mainframe and pe ...
- Annual Planning ... s of both short- and long-term pl ... by major industry classificatio ... .
- Conference ... convenient location in the fall q ...
- Executive ... busy executives to strategic iss ...
- Inquiry Ser ... onal research staff through tel ... s access to INPUT's library of re ...

RESEARCH METH ... rch in computers, communications, a ...

- Research t ... ussions with client representat ...
- Research fo ... with users, vendors, universities ...
- Conclusions ... udgment of INPUT's professional ...
- Professiona ... 20 years of experience in da ... senior management positions with major vendors and users.

For further information on this report or program, please call or write:

INPUT
1943 Landings Drive
Mountain View, CA 94043
(415) 960-3990 Telex 171407

or

INPUT
Park 80 Plaza West One
Saddle Brook, NJ 07662
(201) 368-9471 Telex 134630

INPUT

INFORMATION SYSTEMS PROGRAM

VENDOR WATCH REPORT

LOCAL AREA NETWORKS:
DIRECTIONS AND OPPORTUNITIES

DECEMBER 1983

LOCAL AREA NETWORKS: DIRECTIONS AND OPPORTUNITIES

CONTENTS

LOCAL AREA NETWORKS:
DIRECTIONS AND OPPORTUNITIES

EXHIBITS

I INTRODUCTION

- As part of the INPUT Information Systems Program (ISP), the objective of this report is to assist Information Systems (IS) management to:

  - Understand the current realities of local area network (LAN) capabilities.

  - Determine the appropriateness of current or deferred use of LANs.

- For the purpose of this report, LAN will be defined as an information transmission system that operates over a limited distance (generally less than 3,000 feet) at very high speeds (250 Kbps - 10 Mps).

- This report focuses, unless otherwise specified, on large LANs. Large LANs are those that support bigger hardware, i.e., minicomputers or larger and multiple terminal types. LAN products consist of more sophisticated hardware and software than is found on the PC versions.

- Recently, interest by management information system (MIS) managers and users in obtaining solutions to the complex problem of local delivery systems has intensified. This heightened interest has been accompanied by confusion regarding the myriad of products that are currently available (or proclaimed to be available) in the marketplace.

- The focus of this report is upon the realities of LAN technology and implementation. Actual experience is emphasized; vendor claims are de-emphasized.

- The report has a fourfold purpose:

  - Provide guidance to MIS managers who are responsible for planning local information distribution capabilities.

  - Identify and clarify key issues regarding LAN planning, implementation, and operations.

  - Present an assessment of current and future technology.

  - Provide insight based upon actual user experience.

- The scope of the report includes:

  - Current state of LAN technology.

  - General vendor technical and support capabilities.

  - User experience.

- Specific product evaluations are not included.

- The report is organized into three major areas:

  - Issues.

  - LAN realities and trends.

  - Conclusions and recommendations.

- Comments and questions from clients are invited.

II EXECUTIVE SUMMARY

A. PURPOSE OF THE REPORT

- Local area networks (LANs) have been proclaimed by many to be a panacea for the movement of information within confined physical locations (e.g., single building, floor of a building, building complex). As with all new technologies, the lag time between the "promises" and the realization of these proclamations is generally measured in years.

B. KEY OBSERVATIONS

- The urgent need for comprehensive LAN standards is not being met.
  - Standards are being defined by major vendor products creating a multiplicity of standards.
  - The lack of standards is retarding implementation of LANs that address a broad spectrum of current business needs.
- Currently available network software is inadequate for the business needs of most larger firms.

    - Diversity of protocol support is inadequate.

    - Network maintenance software is generally rudimentary.

    - Supporting software, such as accounting, is generally not available.

    - The ability to interconnect with other networks (i.e., gateways) is just beginning to emerge.

- The total cost of large LAN installations frequently exceeds by a significant amount the sum of the hardware and software components. This realization has been painful for many existing users. Cost elements which may have a very large impact are:

    - Needed software development.

    - Modifications and rearrangements.

    - Installation (building code/union) problems.

    - Required user support.

- Almost all current LAN products are in the process of enhancement. The lack of required software to support user device requirements and network operation/maintenance is recognized by most vendors. This software deficiency is a significant problem for most users.

    - Simple, standalone networks serving a homogeneous set of devices are less impacted.

    - As more complex requirements are addressed, the functional deficiencies become more severe.

- Existing LAN users have discovered that they must be relatively self-sufficient in regard to supporting and maintaining their LANs. Vendor support ranges from being a significant problem to being almost nonexistent.

- The LAN product field is overcrowded with suppliers. About 250 vendors now claim some form of LAN product. Most of these are very small companies with a questionable business survivability.

  - Numerous business failures among LAN suppliers are foreseen in the next three years.

  - The impact of pending AT&T and IBM products will accelerate the demise of many suppliers whose products are noncompatible.

- In general, businesses are selecting LANs in an ad hoc fashion.

  - Most are being obtained by single departments in contrast to some corporate wide procurements.

  - Methodical, objective procurements are the exception.

  - There is a major need to involve qualified corporate expertise in the selection process.

- There is a trend toward the use of twisted pair and fiber optic technology for future LAN offerings.

  - Coaxial cable may be a short-lived technology for generalized LANs.

  - Neither AT&T nor IBM is expected to base their LAN offerings upon coaxial cable; although coax may continue to be supported.

- Fiber optic connectors, which have previously retarded product development, are becoming more flexible and less expensive.

- The user of twisted pair is significant because of the large installed base of this media.

• There has been an overemphasis upon the broadband versus baseband issue. In INPUT's judgment, there are more significant issues to be faced by potential LAN users.

  - What are the current and future requirements?

  - Is a cable-based LAN even appropriate?

  - What is the relationship of the LAN to the PBX/CBX?

  - Will the current LAN be technologically obsolete in two to three years?

• User experience with LAN reliability has been good. The basic technology and associated products work well in a defined, unchanging environment. Availabilities exceeding 99.8% are realizable.

  - Problems with incorporating new features can be severe.

  - Fault isolation is a key problem.

• PBX/CBX technology is progressing rapidly in its ability to satisfy user needs. There is a strong likelihood that this technology will severely impact the marketplace for LANs in the next three to five years, primarily because it is prejustified for voice applications.

C. CONCLUSIONS

- The central theme regarding timing of implementation is WAIT.

  - LAN technology and product development is in a period of great change.

  - The near future of LANs will be heavily influenced by:

    - AT&T announcement.

    - IBM announcement.

    - PBX/CBX development.

  - Very little true, operational experience exists with large LAN products oriented to supporting multiple minicomputers, large hosts, and a variety of terminal types. With few exceptions, users are still at the exceptional stage. The large number of LANs for PCs are very simple and oriented to shared disks and printers, and are considered as being in an entirely different class.

  - The waiting period should be about 12 to 18 months. During that period, announcements will have been made by IBM and AT&T, software will have been improved by major vendors, the role of CBX devices relative to LAN facilities will be better understood and many of the small LAN suppliers will have left the marketplace.

- If a LAN appears to be warranted within the next 6 to 12 months, some key principles are:

- The economic payoff should be projected over a two to three year period, with a strong likelihood that the project will be technologically obsolete after that time.

- Requirements should be carefully outlined and various vendor products methodically assessed against those requirements.

III KEY ISSUES

A. ROLE OF STANDARDS IN LAN CONSIDERATIONS

- From the user perspective there is no disagreement over the urgent need for standards. However, the vested interests of many manufacturers has severely retarded the development of meaningful standards.

- Progress on effective LAN standards is three to five years behind that of comparable wide area communication network standards.

- The basic framework for LAN standards is the International Standards Organization (ISO) Open Systems Interconnection (OSI) model.

- The most significant standards activity has been achieved by the Institute of Electrical and Electronics Engineers (IEEE) 802 committee.

- The initial focus of IEEE 802 was to produce a single, definitive communication standard that would fit within the ISO model. However, this objective proved not to be feasible. In lieu, the "standard" was generalized to recognize the technological differences among the products of several major suppliers.

- The current IEEE 802 standards encompass three distinct areas.

  - 802-3, Carrier Sense Multiple Access/Collision Detection system based upon Ethernet.

- 802-4, Token Passing Bus Network.

- 802-5, Token Passing Ring Network.

- Within each standard, several alternatives exist that reflect differences in areas such as:

  - Data rate.

  - Transmission media.

  - Distance.

  - Modulation/encoding technique.

- IBM support for the 802-5 standard is a significant development.

- AT&T has been conspicuously absent in its support of emerging standards and commitment to them in its own LAN product development. They are adopting the "we are the standard" philosophy.

- Standards activity has been directed toward the lowest levels of the ISO model. However, users are in need of standards at the higher levels which specify agreements for the movement of information between heterogeneous devices. Significant progress on these higher level standards is still several years away.

- In summary, the lack of adequate standards is retarding the meaningful implementation of networks that meet today's business needs.

B. LAN SOFTWARE

- Network software for LANs is rudimentary when contrasted with user needs and expectations. This is a result of the relative newness of LAN technology when contrasted to other, more developed communication product areas.

- The incompleteness of LAN software severely inhibits the user-functional capability. This is particularly true in the area of device interface software. Many LAN products do not support such common protocols as 3270 BSC, 3270 SDLC, and X.25.

- Overall network-support software is incomplete with most contemporary products. Examples of network software that are generally not available or incomplete are:

  - Network accounting.

  - Gateways to other networks.

  - Network control.

  - Access/security.

  - Maintenance tools.

- The small size of many vendors renders it likely that many products will always be lacking the software to provide needed functional capabilities.

- LANs for personal computers have additional software related problems. These include technical issues (e.g., no multitasking, limited or nonexistent file sharing) and licensing issues regarding multiple/shared use of application programs.

C. COST

- A major driving factor for the implementation and use of LANs has been a proclaimed savings in cost over currently used methods. However, actual experience has tended to mitigate these cost savings.

- Some typical planning figures for LANs are:

  - $350 to $700 per device connection.

  - $1 to $2 per foot for cable

  - $5 to $6 per foot for cable installation.

- These planning figures may only represent a small part of the total cost. Potential major cost components that could have a severe impact are:

  - Software development.

  - Higher installation costs caused by union and building code requirements.

  - Modifications and reconfigurations.

- Users have focussed upon connection costs because such costs are directly measurable. It is essential that a broader view be taken.

- Because of the rudimentary system capability of many products, additional software development is normally required. The total cost of such software, whether developed internally or procured, may exceed all other cost components, even exceeding the total of all other costs.

- Many new LAN installations require some form of customization; some will require extensive custom software.

- The combination of union regulations and building code restrictions may result in totally excessive installation costs in certain areas. Some users have determined that these costs outweigh the advantages of a LAN.

- The cost of modifying existing LAN installations precludes or limits the flexibility to relocate devices within a facility. This cost tradeoff effectively reduces LAN flexibility which is generally touted (and believed by potential users) as one of its greatest assets.

D. VENDOR STABILITY AND VIABILITY

- The total number of companies either producing or planning to produce some form of LAN exceeds 250.

- The vast majority of these companies are very small, relatively unknown, and have been in existence for a brief period of time. A high failure rate of such vendors is expected over the next three years.

  - Many of these vendors are overly technological and do not possess the knowledge and/or resources needed to sustain a long-term product-oriented business.

  - A growing user disillusionment with many of the products and support from these smaller vendors will accentuate the trend toward a high failure rate.

- The most significant impact upon the myriad of LAN suppliers will be the entry of IBM and AT&T into the marketplace. Neither has yet formally announced a LAN product; both will announce soon. The impact will be adverse for those vendors whose products cannot be made compatible with the IBM and/or AT&T offerings.

- INPUT strongly believes that the impact on the LAN market made by an IBM LAN product will be analagous to the impact IBM has had on the personal computer market. That is, when IBM announced its personal computer, there were numerous other vendors. Many of those are now "folding" - even a few of the major ones. IBM compatibility seems necessary for survival.

- A commonly heard observation from major companies is "...when IBM has a LAN product, we will install it and test it." This attitude may prove fatal for other, noncompatible products and will be a difficult obstacle for small vendors of compatible products.

- The AT&T offering will gain significant purchaser attention both because of the AT&T stature as well as the key issue of existing in-plant wiring.

- Another greatly significant impact is the emerging development of PBX/CBX products that address local area information movement needs. Some key suppliers are:

  - Northern Telecom.

  - Rolm.

  - Ztel.

  - Mitel.

  - NEC.

- The current offerings of these vendors do not fully meet local distribution requirements; however, in two to three years, overall capabilities are expected to surpass those of competing LANs.

  - In general, the PBX/CBX vendors are larger and more fully capable of satisfying needs and supporting users over the long term.

  - The greater resources combined with a large voice-funded revenue base will enable PBX/CBX vendors to sustain a much more rapid pace of development than most LAN vendors.

- Many of the smaller LAN vendors have enjoyed some initial successes by offering products that address some niche in the marketplace. Users are becoming increasingly wary of these multiple niche approaches and are demanding more ubiquitous solutions to the local distribution problem. The result will be a rapid fall out of many small vendors who are unable to offer a more generalized capability. Typical of the niche approaches are the following:

  - Linking of process control processors with remote terminals.

  - PC LANs to share files and applications among PCs.

  - Shared storage and printers for micro processors and workstations.

- Vendor support is a major issue from the user perspective. Many of the smaller vendors offer little or no installation and post-installation support. Even the larger vendors tend to offer an inadequate level of support relative to the complexity of their products.

- Users are discovering that they must become self-sufficient in regard to LAN support and maintenance. This implies a significantly increased "hidden cost" of LAN implementation. For example, PC networks may be procured at computer stores where support is almost nonexistent.

E. NEW TECHNOLOGY VERSUS PROVEN PRODUCTS

- LAN product selection, like other system decisions, must assess the merits and risks of choosing newly developed (or developing) products that reflect the new technology. Such an approach must be contrasted with the use of more proven products that may lack the fully desired functional capabilities.

- For the entire LAN field, there is a severe lack of proven products. Thus, some risk in new technology must accompany any decision to proceed with a LAN.

  - Currently available products have very restricted capabilities.

  - A methodical approach to LAN product selection is required to avoid user disillusionment.

  - Actual product performance does not meet vendor claims.

- Selecting newer technologies implies:

  - More user involvement.

  - Customization.

  - Scheduled delays.

  - Unforeseen performance problems.

- The risk of choosing "dead-end" technology.

- The risk of selecting a nonstandard product.

• Many companies are using interim approaches while awaiting further LAN developments. The use of a data switch is the most prevalent alternative. While data switches solve some local network problems, they are not considered LANs for the purpose of this report. (Data switch is a digital time division switch using a dedicated circuit per device over limited distances (typically less than 1,000 feet) supporting moderate speed (110 bps - 19.2 Kbps).)

• In summary, for all but very specialized applications, there is no proven product/technology which meets generalized user local distribution needs.

IV LAN REALITIES AND TRENDS

A. MEDIA CONSIDERATIONS

- LANs are implemented on one or more of the following media:

  - Coaxial cable.

  - Fiber optics.

  - Twisted pair.

- Each of these may be further categorized such as thick/thin coaxial cable, shielded/unshielded twisted pair, etc.

1. COAXIAL CABLE

- Coaxial cable has been used for the transmission of information for several decades (e.g., CATV networks). As such, it is not a new technological media. What is new is its adaptation to multistation local networks.

- Principal features of coaxial cable are outlined in Exhibit IV-1.

- The physical problems with coax implementation are proving to be worthy of attention. Retrofit of existing buildings is particularly expensive, with relatively few such installations having been accomplished.

EXHIBIT IV-1

MEDIA CHARACTERISTICS AND COMPARISONS

	MEDIA ALTERNATIVE		
CHARACTERISTIC	COAX	FIBER	TWISTED PAIR
Bandwidth	40-100 MHz	10,000-100,000 MHz	1-4 MHz
Interference Immunity	High	Extremely High	Low*
Installation Ease:			
- Cable	Medium	Good	Excellent
- Connectors	Medium	Difficult	Good
Experience With Media	High	Low	High
Security	Medium	Excellent	Medium
Physical Size	Medium	Small	Small
Exists in Most Facilities	No	No	Yes

*Medium if shielded twisted pair.

- Future trends favor fiber and twisted pair. Fiber represents newer technology; twisted pair represents a large installed base.

  - Coax may be a short-lived technological media for LANs.

  - Physical installation and modification issues will seriously limit growth of coax.

  - Neither IBM nor AT&T is expected to base its new LAN offerings upon coaxial cable.

2. FIBER OPTICS

- Very little experience exists with the use of fiber optics in local network environments. Theoretical expectations and characteristics are high while practical realizations have been quite limited.

- Principal features of fiber optics are outlined in Exhibit IV-1.

- A trend is moving toward the combined use of fiber and twisted pair. Conventional twisted pair would be used for short distances; connections to fiber would be made at some form of concentration.

- The difficulty of connecting to fiber optics cable has retarded the development of products utilizing this technology. New developments are overcoming these difficulties, however.

- Anticipated use by AT&T of fiber optics in LANs will intensify the emerging trend toward use of this media.

  - AT&T has extensive experience with fiber optic technology for long-haul and inter-exchange trunking.

- The combination of fiber optics, as a new technology, with twisted pair, which is omnipotent in all buildings, is a logical AT&T direction.

- Users becoming involved in fiber optic LANs must anticipate considerable involvement, significant customization, and delays in implementation caused by unforeseen problems.

- Two discernible technological trends are:

  - The cost of fiber is decreasing rapidly.

  - Flexibility of interconnection is increasing.

- Both will enhance desirability of incorporating fiber into LAN product offerings.

3. TWISTED PAIR

- Twisted pair represents an old technology. However, it is of great significance because of the extremely large installed base of this media. Almost every building in the U.S. contains some type of twisted pair wiring.

- Principal features of twisted pair are outlined in Exhibit IV-1.

- Ownership of in-plant wiring is confused by the AT&T divestiture. Each of the Bell Operating Companies (BOCs) has an approach for the transfer of in-plant wiring from BOC to building owner. These plans, with their associated costs, become a part of the decision criteria.

- The use of existing twisted pair with newer PBX/CBX devices becomes an alternative approach to LAN implementation.

- This approach offers the most potential for integration among voice, data, and image.

- Because of the dominance of voice in both organizational position and expenditures in most companies, such an approach will often be favored.

- Realizable capacities on twisted pair are increasing. Data rates of 4-10 Mbps are being achieved over modest distances (e.g., 100-1,000 feet). Longer distances seriously degrade realizable capacities.

B. BASEBAND OR BROADBAND?

- Perhaps no issue has occupied attention and coverage in the published literature for LANs as much as the baseband versus broadband debate. This controversy has become the 1980s equivalent of the 1970s "packet versus circuit switching" debate.

- To focus upon this technological comparison is to ignore the broader questions regarding LAN selection and implementation.

  - What are the current and future requirements?

  - Is a cable-based LAN even appropriate?

  - What is the relationship of the LAN to the PBX/CBX?

  - Will the LAN be technologically obsolete in two to three years?

- Within the framework of these broader issues, it is still important to consider the relative merits of broadband and baseband. These are summarized in Exhibit IV-2.

EXHIBIT IV-2

BASEBAND/BROADBAND ISSUES

BASEBAND	
Advantages	Disadvantages
• Modems Not Required • Passive Media • Simpler Connections • Easier Installations	• Digital Only • Limited Distance • Single Channel • Lower Capacity
BROADBAND	
• Analog and Digital • Unlimited Distance • Noise Immunity • Multiple Channels • Well-developed Components (e.g., CATV) • Large Bandwidths • Coexistence of Different Protocols	• More Complex Interface Devices • Less Standardization • More Difficult Design

- Both technologies are being used. The best example of baseband technology is Ethernet, which was originally developed by Xerox. There is no comparable generic example for broadband; several vendors (e.g., Sytek, Ungermann-Bass) have existing broadband products.

- A trend toward hybrid approaches employing combinations of the two technologies is discernible.

C. PROTOCOLS: AVAILABLE AND NEEDED

- Support for a diversity of protocols remains a major problem with most LANs. This is particularly significant since this perceived diversity and flexibility is often presented as the primary reason for a LAN. This lack of generality severely limits the usefulness of current LANs.

- Support generally exists only for teletypewriter (TTY). Some initial announcements of IBM 3270 support have recently been made by a few leading vendors.

- Support is almost nonexistent for:

  - Systems network architecture (SNA).

  - X.25.

  - Inter-network software.

- The developmental cost for new protocol support is high. This large investment combined with the scarcity of experienced programmers will retard the introduction and incorporation of enhanced support capability. For many small vendors, such enhancements will never be realized.

- Software development of protocols for gateways to other (i.e., local, wide area) networks is only just beginning. The general lack of such gateways is a retarding factor to more extensive LAN installations.

- Experience with mixed modes of use, such as interactive and file transfer, on the same LAN have been generally unsatisfactory.

  - Most users have concluded that separate LANs are needed with many installing separate networks.

  - While some vendor products will accommodate both, this should not be an a priori assumption.

D. NETWORK SOFTWARE: SOME, MOST, OR ALL

- The preceding discussions on vendor capabilities and protocols have indicated the reality of a general lack of software availability and support.

- By comparison to familiar technologies, such as mainframe operating systems, intelligent terminals/controllers, and wide area networks, LAN software is in its infancy.

- The impact to the user is manifest in several forms.

  - Need for user-sophisticated programming development and support.

  - Inability to satisfy some requirements.

  - Greater than expected number of problems at installation.

- Overall impact to users is increasing as users begin to move from simple requirements to more complex expectations.

E. RELIABILITY AND MAINTENANCE

- Inherent reliability of the LAN product has been demonstrated to be extremely high. Basic hardware reliabilities are sufficiently high, such that their contribution to problems is almost negligible.

- Availability/reliability problems are most frequently caused by cable-related failures (i.e., physical damage).

- Reliability problems are often experienced when new enhancements to existing products are released. A common user complaint is: "The LAN is OK until new enhancements are introduced; then our operation is disrupted."

- A key issue is the relative unavailability of fault isolation tools. This problem is becoming more acute as LAN size and complexity grows.

  - Often simple tools, such as the ability to "down" a single device or interface unit, do not exist.

  - Another example is a general lack of indicator lights that help determine operational status.

  - There is often no ability to broadcast a message to users (e.g., the network is to be taken down for maintenance).

F. CAPACITY: ACTUAL AND THEORETICAL

- Vendor claims for capacity have not generally been demonstrated. Actual results can be on an order of a factor of two less than vendor-claimed capacities.

- The capacity problem becomes more acute with large numbers of device connections. Very little real experience exists for such configurations.

- However, achievable capacities seem to be sufficiently higher when user requirements are being met.

G. CONNECTION COSTS VERSUS COMPONENT COSTS

- The distinction between connection costs and component costs is an important reality in the area of LANs for PCs.

- LANs for PCs were developed to promote the sale of shared devices, such as disks and printers. They were not developed to provide a generalized local communication capability.

- Some typical costs for connection and component costs on LANs for PCs are:

  - Shared disk system: $2,500 to $6,000.

  - Connection cost: $300 to $500 per connection.

- Therefore, for even modest numbers of devices, the connection costs can completely dominate the system cost.

H. TYPICAL USER EXPERIENCES

- The number of firms actually using LANs, even in a small, isolated environment, remains relatively small. Exhibit IV-3 shows the results of a recent INPUT survey. Only 28% of those responding had actually installed any form of LAN.

- Most companies perceive themselves as being behind in the implementation of LANs. Exhibit IV-4 shows some results from a recent INPUT survey. Almost one-half of the firms considered themselves to be below average in regard to LAN progressiveness; whereas only 8% perceived themselves as being on the leading edge.

- Interviews were conducted with firms representative of business, industry, and educational areas. Over 90% of the organizations interviewed had experienced significant challenges or problems with their installed LAN. The following are representative of the responses:

  - Too few devices supported by the interface (computer manufacturer).

  - Administrative software severely limited (computer manufacturer, aircraft manufacturer).

  - Software release schedules have never been met (computer manufacturer, university, home furnishings manufacturer).

  - Actual number of devices that could be connected and effectively utilized was significantly less than represented (terminal manufacturer, university).

EXHIBIT IV-3

LAN USE

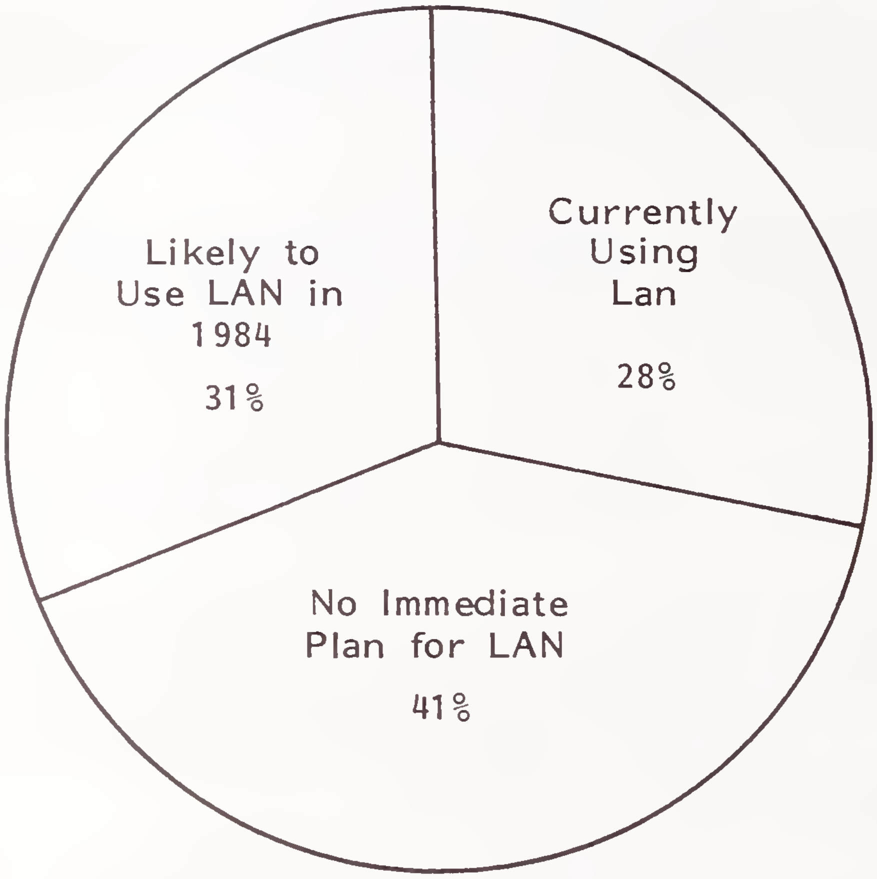
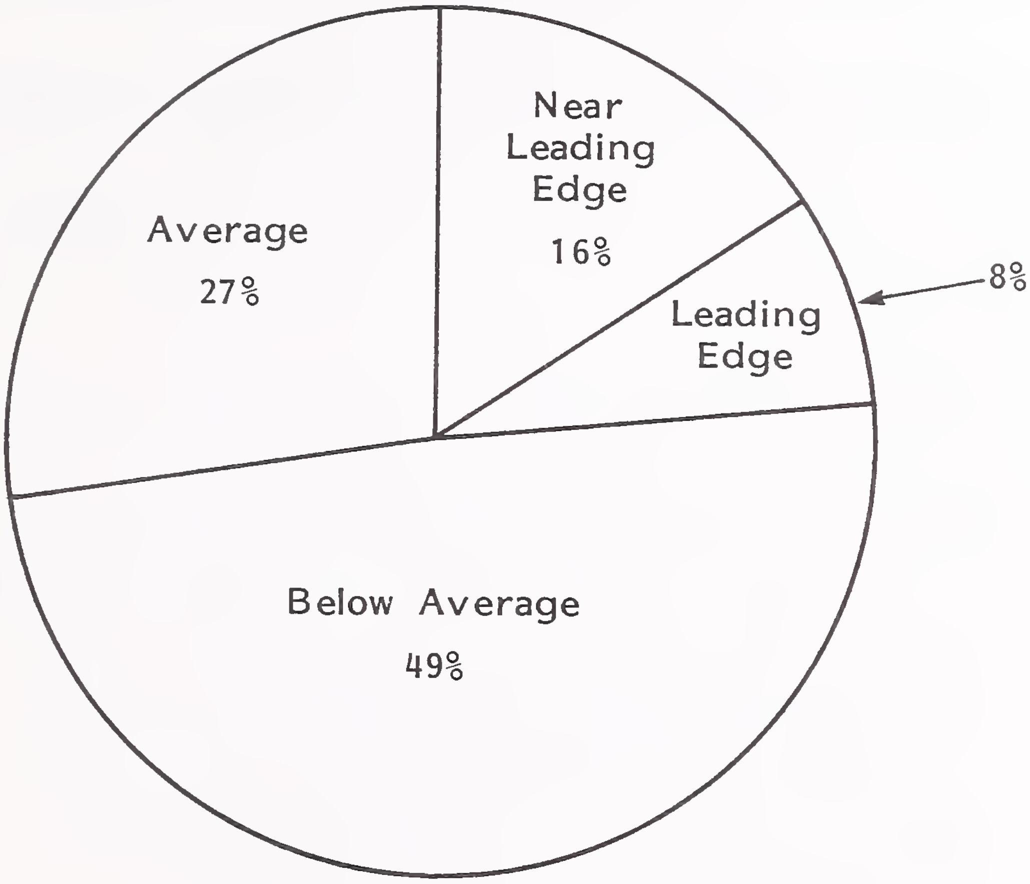


EXHIBIT IV-4

COMPANY PERCEPTIONS OF THEIR LAN PROGRESS



- Significant deterioration in performance when transmitting large file (airline, law enforcement agency).

- File transfer software unreliable (computer manufacturer).

- No facility to broadcast message to connected devices (university).

- Operations manual very poorly written (aircraft manufacturer).

V CONCLUSIONS AND RECOMMENDATIONS

A. TIMING OF IMPLEMENTATION

- The central theme regarding timing of implementation is WAIT.
  - LAN technology and product development is in a period of great change.
  - The near future of LANs will be heavily influenced by:
    - AT&T announcement.
    - IBM announcement.
    - PBX/CBX development.
  - Very little true, operational experience exists with most LAN products. With few exceptions, users are still at the experimental stage.
  - The waiting period should be on the order of 12-18 months. During that period, announcements will have been made by IBM and AT&T, software will have been improved by major vendors, the role of CBX devices relative to LAN facilities will be better understood and many of the small LAN suppliers will have left the marketplace.

- If a LAN appears to be warranted within the next 6-12 months, some key principles are:

  - The economic payoff should be projected to occur over a two to three year period, with a strong likelihood that the project will be technologically obsolete after that time.

  - Do not assume the product can be enhanced to meet future needs.

  - Requirements should be carefully outlined and various vendor products methodically assessed against those requirements.

  - Do not rely on vendor claims; talk to qualified users with actual experience.

B. IN-PLACE OR NEW WIRING

- The ability to utilize existing wiring in a building, to satisfy local data movement needs, is a well-recognized objective.

  - Wiring is expensive.

  - Use of existing wiring favors a PBX/CBX approach.

- Often, major new PBX/CBX installations may require extensive amounts of wiring changes and/or duplicate wiring to effect transition and cut-over. Thus, the apparent advantage of using in-place wiring becomes greatly mitigated.

- Installation of new cabling (i.e., coaxial cable) must be carefully assessed against the following:

    - Technical obsolescence in two to four years.

    - High cost of device and cable configuration changes.

C. PRODUCT SELECTION

- A sample survey of firms who have installed or are procuring local area networks revealed an ad hoc approach to procurement.

    - Many had contracted with a single vendor and were unaware of major alternative suppliers.

    - The general approach was to assess vendor claims, agree those were responsive to needs, and execute a contract.

    - Longer-term needs were not a major factor.

- LANs are primarily being procured for small applications by a department of a firm; corporate-wide procurements are rare.

- There is a strong, almost mandatory, need for potential purchasers of LANs to utilize some qualified corporate technical evaluation resources such as could be found in a centralized information systems organization. Such resources can provide:

    - Objectivity.

    - Technical expertise.

- Structural methodology.

- Key elements in the selection process should be:

  - Adherence to standards.

  - Demonstrated performance.

    - Number of physical devices.

    - Number of simultaneous devices.

    - Reliability.

    - Data capacity.

    - Distance and topology.

  - Functional capabilities.

  - Vendor stability.

  - Total cost.

    - Wiring.

    - Software support.

    - Hardware and software.

    - Configuration changes.

. New feature development.

- Risk of obsolescence.

- Interface to other networks.

- Future requirements.

- Many of the users surveyed as part of this report had serious concerns about locking in to one vendor. In many cases the concern was emphasized because of having made a poor choice of vendor initially.

  - Vendor "lock-in" is a reality; it cannot really be avoided.

  - Therefore, it is mandatory to be more careful in the evaluation and selection process.

D. PRODUCT SOFTWARE PRIORITIES

- Because of the small size and unstructured characteristics of the market, users can have a significant impact upon vendor developmental programs.

- LAN vendor software priorities are oriented to:

  - Expanding the number of protocols supported.

  - Developing gateways to other networks.

- These priorities tend to preclude development of needed maintenance tools, such as fault isolation and performance monitoring.

E. IMPORTANCE OF INTERFACING TO THE OUTSIDE WORLD

- Most currently operational LANs are being used in a small, standalone environment. The most typical application is a small (e.g., 10-15) number of PCs connected to shared devices, such as disk files and printers.

- User requirements are dictating a high priority on LAN interfaces with other entities, such as:

  - Corporate wide area networks.

  - PBX/CBX.

  - Other internal LANs.

  - Vendor networks (e.g., SNA).

- From the user decision viewpoint, it is vital to determine if the potential of LAN requirements is strictly local in nature or must interface with a much wider environment.

  - If the former, the procurement must achieve an acceptable payoff in two to three years with the likelihood that a new approach, with new products, is needed then.

  - If the latter, the best conclusion is to wait. True wide area, local area integration, with all of the necessary support, is not available in a form that satisfies general user needs.

  - A major caution is that there are virtually no truly local requirements for information movement. It is generally only a matter of time before a broader scope of movement is needed.

MANAGEMENT PROGRAMS: Designed for clients with a continuing need for information on a range of subjects in a given area.

- Management Planning Program in Information Systems - Provides managers of large computer/communications facilities with timely and accurate information on developments that affect today's decisions and plans for the future.
- Management Planning Program for the Information Services Industry - Provides market forecasts and business information to information services companies to support planning and product decisions.
- Company Analysis and Monitoring Program for the Information Services Industry - Provides immediate access to detailed information on over 3,000 companies offering software, processing services, integrated systems, and professional services in the U.S. and Canada.
- Management Planning Program in Field Service - Provides senior field service managers in the U.S. and Europe with information and data to support their planning and operational decisions.
- On-Target Marketing - A practical, "how-to" methodology for more effective marketing problem solving and planning. Delivered to clients through workshops and/or consulting services.

MULTICLIENT STUDIES: Research shared by a group of sponsors on topics for which there is a need for in-depth, "one-time" information and analysis. A multiclient study typically has a budget of over $200,000, yet the cost to an individual client is usually less than $30,000. Recent studies specified by clients include:

- Selling Personal Computers to Large Corporations
- Improving the Productivity of Systems and Software Implementation
- User Communication Networks and Needs
- Financial Planning Systems Markets: The Next Five Years

CUSTOM STUDIES: Custom studies are sponsored by a single client on a proprietary basis and are used to answer specific questions or to address unique problems. Fees are based on the extent of the research work. Examples of recent assignments include:

- Organizing for Effective Software Development
- Corporate Plan for Utilizing CAD/CAM
- Annual ADAPSO Survey of the Computer Services Industry
- Analysis of Business Services for a Major Financial Institution
- Study of the Specialty Terminal Market
- Study of Disaster Recovery Services
- Analysis of Software Maintenance Issues
- Review of Software Product Market Opportunities
- Analysis of Network User Requirements

About INPUT

INPUT provides planning information, analysis and recommendations to managers and executives in the information processing industries. Through market research, technology forecasting, and competitive analysis, INPUT supports client management in making informed decisions. Continuing services are provided to users and vendors of computers, communications, and office products and services.

The company carries out continuous and in-depth research. Working closely with clients on important issues, INPUT's staff members analyze and interpret the research data, then develop recommendations and innovative ideas to meet clients' needs. Clients receive reports, presentations, access to data on which analyses are based, and continuous consulting.

Many of INPUT's professional staff members have nearly 20 years' experience in their areas of specialization. Most have held senior management positions in operations, marketing, or planning. This expertise enables INPUT to supply practical solutions to complex business problems.

Formed in 1974, INPUT has become a leading international planning services firm. Clients include over 100 of the world's largest and most technically advanced companies.

OFFICES

Headquarters
1943 Landings Drive
Mountain View, CA 94043
(415) 960-3990
Telex 171407

Detroit
220 E. Huron
Suite 209
Ann Arbor, MI 48104
(313) 971-0667

New York
Park 80 Plaza West-1
Saddle Brook, NJ 07662
(201) 368-9471
Telex 134630

United Kingdom
INPUT, Ltd.
Airwork House
35 Piccadilly
London, W1V 9PB
England
01-439-8985
Telex 23116

AFFILIATES

France
La Nacelle
2, rue Campagne Premiere
75014 Paris
322.56.46
Telex 220064 X5533

Italy
PGP Sistema SRL
20127 Milano
Via Soperga 36
Italy
Milan 284-2850
Telex 310352

Japan
Overseas Data Service Company, Ltd
Shugetsu Building
No 12 - 7 Kita Aoyama
3-Chome Minato-ku
Tokyo, 107
Japan
(03) 400-7090
Telex 26487

Sweden
Athena Konsult
P.O. Persson & Co AB
Box 22114
S-104 22 Stockholm
Sweden
08 52 07 20
Telex 17041

INPUT Planning Services for Management

INPUT

MANAGEMENT PLANNING PROGRAM IN INFORMATION SYSTEMS

VENDOR WATCH REPORT

SELECTING USER FRIENDLY OPERATING SYSTEMS FOR PERSONAL COMPUTERS

JUNE 1983

MANAGEMENT PLANNING PROGRAM IN INFORMATION SYSTEMS

OBJECTIVE: To provide managers of large computer and communications systems with timely and accurate information on developments that affect today's decisions and plans for the future.

DESCRIPTION: Clients of this program receive the following services each year:

- Impact/Planning Support Studies - In-depth reports dealing with the impact on users of projected managerial, personnel, and technological developments over the next five years. Studies include analyses and recommendations.
- Technology and Management Issue Briefs - Analyses of the probable moves of major computer/communications vendors in operating systems, data base/data communications, software, mainframes, office systems, value-added networks, personal computers, and other strategic areas.
- Residual Value Forecasts - Five-year forecasts of the residual values of major mainframe and peripheral equipment.
- Annual Planning Report - Analyses and composite forecasts of both short- and long-term plans of computer/communications managers by major industry classification. Published during the second half of the year.
- Conference - National conference for all clients held at a convenient location in the fall quarter.
- Executive Bulletin - A quarterly newsletter alerting busy executives to strategic issues that could affect their decision making.
- Inquiry Service - Individual consultation with professional research staff through telephone inquiries and visits as needed. Includes access to INPUT's library of research materials.

RESEARCH METHOD: INPUT carries out extensive research in computers, communications, and associated fields:

- Research topics are selected by INPUT based on discussions with client representatives.
- Research for this program includes professional interviews with users, vendors, universities, industry associations, and other analysts.
- Conclusions derived from the research are based on the judgment of INPUT's professional staff.
- Professional staff supporting this program average nearly 20 years of experience in data processing and communications, including senior management positions with major vendors and users.

For further information on this report or program, please call or write:

INPUT
1943 Landings Drive
Mountain View, CA 94043
(415) 960-3990 Telex 171407

or

INPUT
Park 80 Plaza West One
Saddle Brook, NJ 07662
(201) 368-9471 Telex 134630

INPUT

INFORMATION SYSTEMS PROGRAM

VENDOR WATCH REPORT

SELECTING USER FRIENDLY OPERATING SYSTEMS FOR PERSONAL COMPUTERS

JUNE 1983

U-V08
1983

AUTHOR

TITLE

SELECTING USERFRIENDLY OPERATING
SYSTEMS FOR PERSONAL COMPUTERS

DATE LOANED	BORROWER'S NAME
4.11	Jack

SELECTING USER FRIENDLY OPERATING SYSTEMS
FOR PERSONAL COMPUTERS

CONTENTS

SELECTING USER FRIENDLY OPERATING SYSTEMS FOR PERSONAL COMPUTERS

EXHIBITS

I INTRODUCTION

- This report is issued as part of INPUT's Information Systems Program (Software Planning Module).

A. IMPORTANCE OF THE OPERATING SYSTEM DECISION

- It is no longer possible for Information Systems (IS) planners and managers to ignore the personal computer invasion.

  - Users are already looking wistfully at the high end of the personal computer spectrum for powerful features previously available only on minicomputers or mainframes.

  - Now the second generation of personal computers is arriving in force. It is characterized by 16-bit microprocessors more than twice as fast as the first generation, which can address megabytes of memory, rather than the 64K of the 8-bit machines. Some machines also offer multitasking, packaged software written in high-level languages, large-scale data base manipulation, sophisticated data communication and other advanced capabilities. End users are once more enticed by the promise of freedom from the IS organization.

- Operating systems written for the first generation of personal computers provide the bare minimum of support facilities: I/O management, file management, primitive or no interrupt processing, perhaps a rudimentary job stream feature, and a limited set of utilities. Memory management, protection, or multitasking facilities are usually not provided. They are not really necessary in an 8-bit, single-user environment.

- Second generation personal computer operating systems consist of extensions and expansions of first-generation systems, downsizing of minicomputer-based operating systems, and a few new entries. Many of these claim to be user friendly, and some may even deserve the designation.

- But none of the personal computer operating systems serves all categories of users well, nor do they replace mainframe operating systems in power or flexibility. To date, vendors have given little attention to integrating personal computer operating systems into the overall IS environment, which complicates the IS planner's decision: choose now or wait (at the risk of having the user present the planner with another de facto choice)?

B. REPORT SCOPE

- This report analyzes five categories of so-called "user friendly" personal computer operating systems that are available on a variety of vendor machines:

  - UNIX and its variants and look-alikes.

  - CP/M and its variants.

  - MS-DOS (PC-DOS) and its variants.

- UCSD p-System.

- Others.

- Apple's SOS, Commodore's DOS, and Tandy's TRS-DOS are not considered in this report because they are proprietary to those single machines, but each of these manufacturer's personal computers can utilize one or more of the five categories of user friendly operating systems (with the possible requirement of an additional circuit board).

- The report will also discuss the characteristics of a user friendly operating system, assess current offerings according to those characteristics, describe their historical and likely future development, and present recommendations for the IS planner.

- Interested readers are referred to two of INPUT's earlier reports:

  - New Directions In Operating Systems, Data Base Management, And Communications discusses the IBM System/38 operating system, an implementation of many user friendly characteristics on a larger machine.

  - Personal Computers In The Information Systems Strategy describes the roles IS departments may assume to help manage the personal computer explosion within their organizations.

II EXECUTIVE SUMMARY

- Information systems (IS) planners need to consider the role of a personal-computer-based, user friendly, operating system in relation to overall systems software strategy.

- The migration from standalone 8-bit personal computers to 16- and 32-bit multitasking, multiuser models accessing mainframe data bases presents the IS planner with difficult decisions: which personal computer operating systems should be recommended as the corporate standard? And are there any exceptions?

  - Operating systems commonly considered user friendly (such as UNIX) may offer powerful features for professional software developers but have an unfriendly effect on nonprogrammers.

  - Similarly, menu-oriented, "bulletproof" operating systems may be too tedious, inflexible, and primitive for heavy production or software development.

- The compromise approach, to choose a relatively user friendly system like MS-DOS (PC-DOS) that is widely used, has a fair amount of packaged software available, and offers an upward migration path to XENIX (a UNIX implementation by Microsoft), will best suit the combined single user/multiuser organization unless software development is intended as a primary application.

- For the organization that already has a large investment in a CP/M-based machine, upgrading either to MP/M or Concurrent CP/M permits taking advantage of the large quantity of packaged software already available. They may use one of the "shell" front-end programs to make CP/M friendlier or switch to CP/M+.

  . Apple's operating system upgrade strategy is not clear, but LISA can use XENIX, and other Apple computer models can use CP/M with the additions of Softcard (a hardware board).

  . Digital Research will soon be offering a compability path to UNIX-based software but does not have one now. All its new software is being written in the C language, however.

- Special circumstances would make native UNIX or one of the lesser known alternatives (p-System, OASIS, PICK, etc.) preferable.

  . For UNIX it could be an existing investment (or the ability to make such an investment) in DEC gear, plus a heavy organizational emphasis on software development and a commitment to the C language.

  . For the p-System, it could be an overriding language commitment to UCSD PASCAL (p-System) or the need for application portability within an environment with mixed vendor PCs.

  . Selection of other operating systems would need to be based on the unique advantages they would offer to a particular organization. This approach requires a willingness to tolerate a lover level of support, less packaged software, and the expense and aggravation of "going it alone."

- INPUT recommends that IS planners:

  - Include specific personal computer operating systems in their strategic plans and corporate standards.

  - Use decision matrices such as those in Exhibits II-1, III-2 and III-3 to help select appropriate candidates. Exhibit II-1 presents INPUT's recommended operating system choices (column G), assuming an organization possesses characteristics listed in columns B through F. These decision matrices are presented primarily from the point of view of the nonprogrammer or less experienced programmer and represent ratings averaged across all versions of each system.

  - Remember that native UNIX is a model, not a standard. There are many situations it does not fit very well.

  - Consider the use of various "bridge" programs if it becomes necessary to use a different operating sytem than the one already in place.

  - Inform those users who require it that multiple processor boards (i.e., from Sritek or others) can provide the versatility of all the operating systems at a lower cost than multiple personal computers.

  - Recognize that application software vendors have more to do with the choice of a personal computer operating system than the IS organization does. Avoid the losers and dead ends, even if they are technically elegant.

PERSONAL COMPUTER OPERATING SYSTEM DECISION MATRIX

COLUMN #: A	B	C	D	E	F	G	H
			INTENDED USE				
ALTERNATIVE NUMBER	EXISTING PERSONAL COMPUTER INVESTMENT	ANTI-IBM BIAS?	REPLACING TIME-SHARING?	INTEGRATING WORD PROCESSING?	DISTRIBUTED DATA PROCESSING?	SUGGESTED CHOICE	REASONS/CAVEATS
1.	None or Minor	No	Yes	No	No	PC-DOS	1,2
2.	None or Minor	No	No	Yes	No	p-System	A
3.	None or Minor	No	Yes	Yes	No	CP/M	1, 2, 3
4.	None or Minor	No	No	No	Yes	XENIX	4, C
5.	None or Minor	—	Yes	Yes	Yes	None	B
6.	None or Minor	Yes	Yes	Yes	No	CP/M	1, 2, 3
7.	None or Minor	Yes	No	No	Yes	Any UNIX	4, C
8.	IBM-PC or 8086/8088*	No	—	—	—	Same as 1-5 above	Same as 1-5 above
9.	Z80 or 8080/8085**	—	Yes	No	No	CP/M	1, 2, 3
10.	Z80 or 8080/8085**	Yes	No	Yes	No	CP/M	1, 2, 3
11.	Z80 or 8080/8085**	No	No	Yes	No	p-System	A, D
12.	Z80 or 8080/8085**	—	Yes	Yes	No	CP/M	1, 2, 3
13.	Z80 or 8080/8085**	—	No	No	Yes	None	B, E
14.	Z80 or 8080/8085**	—	Yes	Yes	Yes	None	E
15.	68000***	—	Yes	Yes	Yes	Any UNIX	3, D
16.	Any Other Single Type of PC	—	—	—	—	?	B
17.	Combinations of Above Types	—	—	—	No	p-System	5

Reasons:
1 = Many application packages.
2 = Strong upgrade path (but requires different operating system versions).
3 = More versatility.
4 = Multiuser capability.
5 = Wide ranging portability.

Caveats:
A = If IBM Display Writer, else CP/M.
B = No single solution.
C = If C language, else pioneer.
D = Not many application packages.
E = Not powerful enough.

Key: — = Indifferent, CP/M, UNIX = Family of operating system.
*Includes Burroughs, Convergent,Technology, DEC, Eagle, NCR.
**Includes Osborne, TRS-80, North Star, Xerox, Heath/Zenith, and Wang.
***Includes Altos, Charles River, Sage, Fortune, Wicat, Codata, and other integrated systems.

III CHARACTERISTICS OF PERSONAL COMPUTER OPERATING SYSTEMS

A. CONCEPTS AND FACILITIES OF AN OPERATING SYSTEM

- The function of an operating system is to decouple the user and the application software from the hardware. This requires at least three facilities:

  - Control and management of the computer's resources, including:

    - Memory allocation.

    - Disk file allocation.

    - Command processing.

    - Job stream management.

    - Internal timing requirements (priorities).

    - Management of the user (operator) interface.

    - Recognition and management of hardware error conditions.

  - Residence and operation of primitive programs that interact with and control the unique hardware components. For example:

- Getting a character from the keyboard.

- Putting a character on the cathode-ray tube (CRT) screen.

- Converting data from one code representation to another.

- Handling timing and access sequences of individual peripheral devices.

- Support and operation of the system utilities, used to:

  - Copy, catalog, reorganize, or delete data and/or program files.

  - Create and edit programs and data.

  - Display the status of various system capacities, e.g., disk file space available.

  - Customize the operating system to various device and size configurations.

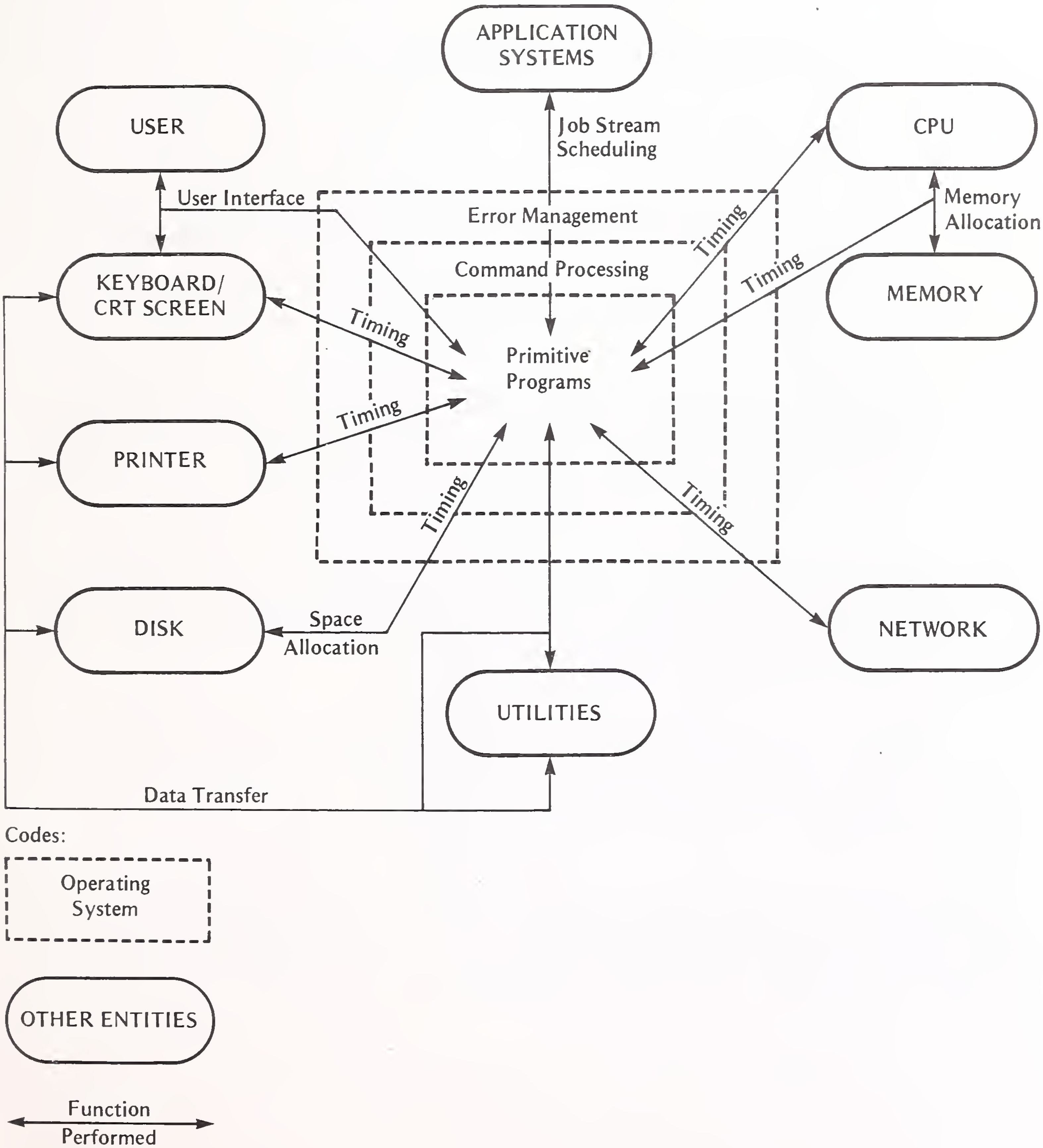
  - Optionally, interface with one or more program development assemblers, compilers, and debugging facilities.

- Exhibit III-1 graphically portrays the relationship of the operating system and the entities with which it interacts.

• By itself, however, an operating system is not worth much. The high-level application languages it supports, the migration path it provides, and the portability, size, speed, and stability of applications developed under it deserve at least as much attention as the operating system itself.

EXHIBIT III-1

INTERACTION BETWEEN OPERATING SYSTEM
AND OTHER ENTITIES



- It is also important to keep in mind that features of certain operating systems can make it easier for application developers to develop more effective software. For example, virtual memory management helps developers create more user friendly programs.

B. CHARACTERISTICS OF USER FRIENDLINESS

- User friendliness depends upon the category of user for whom the system is designed.

  - Programmers value terseness, power, efficiency, and elegance. They are seldom put off by cryptic messages and commands (although even they may be upset if they lose several hours of work because a "delete" command was accidentally invoked).

  - End users, by way of contrast, may require detailed messages or even several levels of messages from the system, depending on how familiar they have become with its responses and characteristics.

  - Selection of responses from a menu is fine while the user is learning the system, but it soon becomes tedious and even patronizing once the range of legitimate responses is known.

  - Both end users and programmers suffer from powerful commands that inflict strong negative effects because they are insufficiently safeguarded.

- Understanding the distinction between the needs of programmers and nonprogrammers is critical for choosing a user friendly personal computer operating system. So are such compatibility questions as:

- Will this be a multitasking or a multiuser system?

- Which high-level languages and/or software packages will be run on this machine?

- Are there special requirements for interfacing, such as to a particular local area network?

- Is there a real time (i.e., clock interrupt) requirement?

• Each of these factors could have an overriding influence on the choice of an operating system, as could the need to match a specific brand of personal computer that utilizes only the manufacturer's proprietary operating system.

• Characteristics of user friendliness include:

  - Consistency of approach, e.g., operating parameters always stated in the same sequence.

  - Simplicity and understandability.

  - Optimization of the normal situation rather than the rare exception, e.g., prestating the normal response.

  - Accessibility of a "help" function.

  - Protection from unanticipated negative results, e.g., providing an "undo" (oops!) function key.

  - Grouping of similar functions, e.g., providing all catalog mangement functions in a single program.

  - Ability to vary responses to match user skill levels.

- Tolerance of user errors, e.g., providing a second chance or bypass, rather than returning to the operating system.

- Incorporation of an integrated systems development "toolbox" of substantial power and flexibility.

- Ability to provide application consistency across different personal computers (i.e., minimum program changes required when transferring to a new machine).

No current personal computer operating system (or mainframe operating system, for that matter) furnishes all of these capabilities, although some come close. Vendors of personal computer operating systems have largely ignored the needs of end users who wish to interact directly with the operating system. This is because vendors primarily sell to OEMs who have experienced programmers on their staff. Each IS organization will have to rank the capabilities it values according to the expected classes of its users (nonprogrammers up through experienced programmers). Other characteristics may have to be added to the list depending on individual company circumstances.

- Exhibit III-2 presents INPUT's rating of four popular personal computer operating systems according to the above 10 user friendly characteristics. This point of view tends toward the nonprogrammer and less experienced programmer end of the scale. Ratings are averaged across all versions of each operating system family. Summaries are shown for both end users and IS professionals.

- Besides the characteristics of user friendliness, it will also be necessary for the IS planner to consider:

EXHIBIT III-2

INPUT'S RATING OF USER FRIENDLY CHARACTERISTICS

ID #	CHARACTERISTIC	UNIX	CP/M	MS-DOS	p-SYSTEM
1.	Consistency of Approach	Poor	Fair	Good	Good
2.	Simplicity/Understandability	Poor	Fair	Good	Fair
3.	Optimizes Normal Case	Good	Fair	Good	Fair
4.	Built-in Help Function	Fair	Fair-Poor*	Fair-Poor*	Very Poor
5.	Protects from Harmful Effects	Very Poor	Poor	Fair	Poor
6.	Groups Similar Functions	Fair	Fair	Good	Good
7.	Varies Response by Skill Level	Poor	Fair*	Fair*	Poor
8.	Tolerant of User Error	Poor	Poor	Fair	Fair
9.	Has Development Tool Box	Very Good	Fair	Poor	Fair
10.	Application Consistency across Different Machines	Fair	Fair	Poor	Good
RATINGS SUMMARY (END USER)		LOW	LOW	MEDIUM	MEDIUM
RATINGS SUMMARY (I.S. PROFESSIONAL)		MEDIUM - HIGH	LOW - MEDIUM	LOW	MEDIUM

*Depends on version.

- Application software availability.

- Physical (compatibility) constraints.

- Support.

- Price.

- Any other factors that would normally be part of a systems software decision.

• INPUT's overall rating of suitability of the four leading personal operating system contenders by the factors INPUT considers essential are shown in Exhibit III-3.

EXHIBIT III-3

INPUT'S RATING OF SUITABILITY OF FOUR PERSONAL COMPUTER OPERATING SYSTEMS

CHARACTERISTIC	UNIX FAMILY	CP/M FAMILY	MS-DOS	p-SYSTEM
Overall User Friendliness Rating	Low	Low	Medium	Medium
Suitable for:				
Office Worker	Poor	Maybe	Yes	Yes
Mid-Management/Professional	Maybe	Yes	Yes	Yes
Programmer	Yes	Yes	Extra Dollars	Yes
High-Level Languages Available	Very Few	Few-Much***	Some	Few
Multitasking System	Yes	Some Versions	No	Yes
Multiuser System	Yes	Some Versions	No	No
Security	Fair	Fair	Poor	Poor
Network Interface	Yes	Yes	No	No
Timer Interrupt Facility	Limited	Some Versions	Yes	No
Portability	High	Some	Low	Very High
Migration Path	Yes	Yes	Yes	No
Hardware				
Z80/8080/8085	Yes	Native	Yes	Yes
8086/8088	Yes	Yes	Native	Yes
Z8000	Yes	No	Yes	Yes
68000	Yes	Yes	Yes	Yes
PDP-11	Native	No	No	Native
VAX	Yes	No	No	No
Others	Yes	No	No	Yes
Memory Size Requirement	Large	Small	Small	Medium
Support				
By Manufacturer	Extra Dollars	Yes	No	Yes
By OEM	Yes	Yes	Yes	Yes
By User Group	Yes	Yes	No	Yes
Price**	M-VH	L-M	L	L-H

*Small = Runs in 64K
Medium = Runs in 128K
Large = Runs in 256K
(Some versions of each are larger.)

**L = Low (less than $100)
M = Medium ($100-$500)
H = High ($500-$1500)
VH = Very High (more than $1500)

***CP/M-68K has few, CP/M-86 has many, and CP/M-80 has much.

IV COMPARATIVE ASSESSMENT OF PERSONAL COMPUTER OPERATING SYSTEMS

- This chapter discusses the historical development, present level of acceptance, degree of user friendliness and other facilities, drawbacks, related products (if any), and likely future development of UNIX, CP/M, MS-DOS, and the UCSD p-System. Also included are some brief comments concerning OASIS and PICK.

A. UNIX, ITS VARIANTS AND LOOK-ALIKES

- INPUT has issued several reports describing or referencing UNIX as a powerful development language, the basis for "Programmer's Workbench" (a productivity tool), and a new concept in programming. In fact it is the latter characteristic that has the broadest implications for planners. Much of the power and productivity of UNIX come from the concept that impelled its development: a thoroughgoing effort to reduce its component functions to the atomic level and to provide a means to assemble them consistently and easily to perform larger tasks.

- This combination of general design, device independence, and library of predefined development tools, plus the decided advantage of being written in one of the more portable source languages, makes UNIX a formidable contender for the leading 16-bit operating system of the mid-1980s.

HISTORY

In the late 1960s, GE developed a multiuser interactive operating system known as MULTICS. This system was (and is) known for its high level of data security and was therefore not suitable for development projects being undertaken by Bell Laboratories' teams of programmers who needed to freely share data files and have access to each other's work.

Ken Thompson, one of the scientist/programmers at Bell Labs, created a transportable language called "B" and used it to develop an operating system, an assembler, and a set of utilities for the PDP-7 minicomputer available to the team. He called the package UNIX, in obvious contrast to MULTICS.

- Dennis Ritchie, another Bell Labs scientist, revised the language (incrementing its title to "C") and rewrote the UNIX software in the new language.

- Other Bell Labs people associated with the development of UNIX included Brian Kernighan and P.J. Plauger. They are more widely known than Thompson or Ritchie because of their collaboration on two books, The Elements of Programming Style and Software Tools.

UNIX has been a "cult" system for many years based on its restricted use within the Bell community, academia, and the ARPA network.

- Unquestionably this situation was aided by Bell's narrow licensing policy, which offered the system at cost (a few hundred dollars) to public agencies but charged commercial users some $28,000 for a license that provided only "as is" source code.

- By 1978 UNIX installations nevertheless exceeded 600, and the system had gone through six versions. AT&T then altered its licensing policy to permit binary sublicensing (object code only) at $2,500. UNIX now began to attract imitators and resellers who cleaned up some bugs and quirks, added support and documentation, ported it to other computers besides DEC machines (including micros), and its popularity began to surge.

- Version six was followed in 1979 by version seven, which in turn was substantially enhanced by the University of California at Berkeley and known as "Berkeley UNIX" or V7BSD or V32. This version provides virtual memory, LISP, PASCAL, a screen editor, and network support. Digital Equipment Corporation has announced support for this version.

- Meanwhile, in 1981, Bell issued another version improved from version six, but less powerful than V7BSD. Bell called this version System III, and it is the basis for many of the clones. In January 1983 still another version, known as System V, was announced. System V features a kernel whose specifications are "frozen" for all future generations. Bell will itself provide support for this system, which is at parity with the version currently in use within AT&T. The support costs extra, however, and covers only VAX and PDP-11/70 implementations, and only for full source code licensees ($43,000 minimum).

2. ACCEPTANCE

- UNIX has many variants and look-alikes operating on all levels and types of computers from personals to mainframes. A sampling of UNIX clones is shown in Exhibit IV-1, together with selected attributes of each.

  - INPUT estimates that there are now more than 10,000 installations of UNIX or its counterparts. Half of these are XENIX installations. But the future for the UNIX clones is clouded by Bell's decision to support the product itself and the likelihood it will offer a workstation of its own using Western Electric's 3B20 minicomputer or 3B5 micro.

CHARACTERISTICS OF SELECTED UNIX CLONES

NAME	VENDOR	APPROXIMATE DOLLARS	VERSIONS EXIST FOR*			MINIMUM MEMORY	BASED ON VERSION	SUPPORTS				
			8086	Z8000	68000			C	PASCAL	FORTRAN	BASIC	COBOL
COHERENT	Mark Williams Co. Chicago	$500/IBM-PC More/Others	X	X	X	128K	Prop.	Y	$	$	N	N
IDRIS	Whitesmiths, Ltd. Concord, MA	$2,000	X		X	128K	Prop.	$	$	$	N	N
UNOS	Charles River Data Systems Natick, MA	$3,000			X	256K	Prop.	$	$	$	$	O
XENIX	Microsoft Bellevue, WA	**	X	X	X	192K	V7B	Y	$	$	$	$
ZEUS	ZILOG Campbell, CA	**		X		256K	V7	Y	$	$	$	N

Codes: *
**Available only through OEMs.
$ = Extra cost.
O = Available form other vendors.
Prop. = Proprietary, not a Bell licensee.
Y = Yes, Included.
N = Not offered.

- In any case there is little doubt that the number of UNIX and UNIX-based installations will increase dramatically over the next three to five years. Tandy's selection of XENIX as the standard operating system on their Model 16 will alone potentially double the number of XENIX users.

- In itself, this growth in popularity is not a sufficient reason to jump on the UNIX bandwagon. As stated earlier, it depends on the intended use and the intended user.

3. IS UNIX USER FRIENDLY?

- Most users would say no.

- UNIX was developed for professional programmers, i.e., those writing software for others to use. UNIX is a powerful system, but it lacks most of the features that would make it attractive to casual users.

  - UNIX has an extensive command list that is anything but intuitive. At last count, the pocket reference guide in use at Bell Labs had 14 pages.

  - The mnemonics of some commands are notoriously obscure. When they are strung together (one of the powerful features of the language), a liberal use of comments is in order to describe the net result. In this respect, UNIX is not unlike APL, C, or assembly language, except that its functions have more power.

  - Despite improvements in later versions, it is still possible to incur disastrous effects from what amount to typos in well-intentioned commands. Warning messages are seldom given.

- On the other hand, the system provides a wealth of support for professional programmers. These tools are applicable, of course, only to the languages for which a UNIX compiler or assembler is available.

- These include C, FORTRAN 77, RATFOR, and for some of the variants and look-alikes, LISP, PASCAL, BASIC, MICRO-COBOL, and Ryan McFarland COBOL.

- They do not include PL/I or ANSI COBOL, the most common business system languages.

• UNIX has extensive text-editing facilities, again with many powerful features, but they are not word processing facilities.

  - The editor and the formatter are separate programs requiring any changes to be processed first through the editor then through the formatter and so on, an awkward procedure at best.

  - However, the UNIX system contained an automatic spelling checker long before any of the word processors did.

• UNIX also offers office automation facilities in the form of a built-in electronic mail system and a flexible, hierarchical filing system. These capabilities do not presently exist in any of the other operating systems discussed in this report. That they are an integrated part of the UNIX system is a decided plus.

4. THE UNIX CLONES

• Despite its alleged universality, UNIX (even in its native versions) is not fully compatible from version to version. It is, however, reasonably portable from one machine to another, since there are C compilers available for many different machines.

• UNIX clones, therefore, come in two major varieties: those that are licensed from Bell and are subsets of one of the versions of UNIX (usually V7 or System

III), and those that were written independently to conform to the UNIX manual. Altogether there are about 35 separate implementations of UNIX, covering micros to mainframes.

- Chief among the licensed versions for personal computers is Microsoft's XENIX.

  - Microsoft makes this product available to OEMs rather than to retail outlets. It has been implemented on 8086, Z8000, and 68000 processors.

  - It is one of the fuller versions, although it does not presently support virtual memory, nor does it provide a high-level debugger. There are PASCAL, FORTRAN, BASIC, and COBOL (but not ANSI COBOL) compilers available from Microsoft.

- Among the unlicensed (i.e., independently written) versions, three are of particular interest: UNOS, IDRIS, and COHERENT.

- UNOS is probably the most widely known. It is a product of Charles River Data Systems, manufacturer of the Universe 68 supermicro.

  - The UNOS system is based on and compatible with UNIX V7 but extends its capabilities to provide English names for the functions, shared data, stronger disk management facilities (bit-mapped allocation, bad-block avoidance), and real time control of system scheduling, memory, and devices.

  - UNOS does not provide the UNIX compiler writing aids or the typesetting facilities, but it does provide a DBMS with a dictionary, keyed and generic keyed access, and relative access. It also includes transaction stamping, logging, and back-out facilities.

- The Universe 68/UNOS product is intended for the OEM market.

- IDRIS and COHERENT, on the other hand, are sold directly to the end user, but neither is available for the 8-bit processors (which could only run "baby" UNIX systems anyway).

  - IDRIS was the first UNIX-like system to be developed. It is a product of Whitesmiths, Ltd. (whose president is P.J. Plauger) and was specifically rewritten from scratch to provide support for the smaller user who couldn't afford the Bell license and didn't have a PDP-11 to run it on. Whitesmiths was founded to produce a C compiler, and IDRIS was a logical product to follow it.

  - IDRIS is based on UNIX V6 and does not offer all of the facilities of some of the larger implementations, but this could be an advantage to the user who has only smaller machines. IDRIS can run on a 128K 68000-based machine with a minimum 5 MB hard disk.

- COHERENT, from Mark Williams Company, is the lowest priced, reasonably sized UNIX look-alike. It is available for the IBM-PC at $500, including a C compiler; PASCAL and FORTRAN compilers are also available at extra cost. This is a low-cost, low-risk way for IS planners to gain first-hand familiarity with C and UNIX. It may not be the system to stay with if they don't have an IBM-PC since it costs $2,000 and up for other machines. However, it is available for the Z8000 and 68000 processors, as well as the 8086 and the PDP-11.

5. DRAWBACKS

- In addition to the relative lack of user friendliness, it ought to be noted that UNIX is a large operating system. For any of the fuller implementation, at least a quarter megabyte of main memory is required, with five or preferably ten megabytes of hard disk storage available. The full system (including utilities) approaches nine megabytes of code. This is not to say the system cannot be crammed into something smaller, but neither it nor the user will be happy with the results.

- Until recently the support situation was definitely a barrier. It promises to be better with DEC and Bell supporting the product, but their involvement at present is still minor.

- UNIX is quite an expensive proposition compared to some of the other operating systems discussed in this report.

  - Depending on the specific implementation, a few of the clones (such as COHERENT on the IBM-PC and XENIX on the Tandy Model 16) are less expensive.

  - Since there is a great variation in both price and capability, each of the alternatives needs to be evaluated separately.

- UNIX does not have a native DBMS associated with it. It is oriented primarily to DEC hardware and DEC-related software. This is a distinct disadvantage for the IBM-oriented organization (or Honeywell or Univac, by the same token).

- UNIX is available from Amdahl on their machines under the name UTS (Universal Timesharing System). It operates as a virtual machine under VM/370. Amdahl requires that the user also hold a license from Western Electric and pay a $3,000 monthly license fee to Amdahl.

- IBM also offers a variant, CPIX, on the Series/I and has announced it will offer another variant on the 4300 series (which means it could run on the 30XX series as well).

- But the most important drawback of all is the almost total absence of software packages written to run under UNIX. This situation will improve in the next few years, but it is currently a serious obstacle when choosing an operating system that is friendly to end users. (The least user friendly system is the one that makes you write your own software.)

6. FUTURE DEVELOPMENTS

- UNIX already spans the range from personals to mainframes, from 16-bit processors through to 32-bit. It is written entirely in C, and can be ported to any future machine for which there is a C compiler.

- There are now over 30 companies offering UNIX or a UNIX-like product, including two giants, Bell and DEC. UNIX is widely used within Bell. It is reasonable to expect that UNIX will be around for a while and that its capabilities will be improved.

- Major weaknesses in the system (although they were considered to be design strengths) are the lack of file and record locks, which permits data to go out of synch, and the limited real time capabilities. (UNIX has a clock, but the library of jobs to be run is not scanned very frequently so, in effect, the clock is only accurate to between 10 minutes and an hour.)

    - For some users these are drawbacks; for others they are advantages because changing them will inevitably add more complexity and cost to the system.

    - In any case, they have been or are being addressed by some vendors, but not by Bell.

- How many clone vendors will survive is an important issue since some of them were motivated to enter the market by the absence of Bell support for the UNIX product.

- INPUT believes that XENIX, UNOS, and some of the other look-alikes that have chosen particular market niches will survive. Those whose only advantage is low cost, or end-user support, or operation on a minimal configuration will be vulnerable.

B. CP/M AND ITS VARIANTS

- By far the most widely used, and next to UNIX the oldest, of the personal computers operating systems discussed in this report is CP/M. It has grown into a whole family of operating systems and has been the inspiration (by necessity, some say) for a number of similar products.

- It, too, has many of the characteristics of a "cult" system but for a different reason: its users have the "esprit de corps" of survivors who together have learned to live with its handicaps.

1. HISTORY

- Gary Kildall, the father of CP/M, was working as a consultant to Intel in 1974 to develop a high-level language known as PL/M (Programming Language for Microcomputers) for their new 8080 chip. The state of the I/O art for micros in 1973 called for the use of punched paper tape on a teletype machine, a time-consuming process. Having access to a Shugart eight-inch disk drive, Kildall used PL/M to write a disk operating system, which he called CP/M (Control Program/Monitor).

- Intel decided it was not interested in CP/M, so Kildall formed Digital Research Inc. in 1976. He licensed his operating system to several companies who were offering floppy disk products and found it advantageous to use his system rather than develop their own.

- The companies included Tarbell Electronics, who included it free with their floppy disk controller kit. Hobbyists were drawn to it like bees to honey.

- Imsai, another of the early microcomputer manufacturers, provided IMDOS (CP/M in disguise) for their popular machine when it became apparent their own operating system could not be delivered on time.

- These systems all used the S-100 bus to interface peripherals, and this interface soon became a standard (despite technical deficiencies) because it was widely available. The fortunes of CP/M paralleled this course, for much the same reason and despite many of the same objections.

- Lifeboat Associates, one of the first software-only stores, was formed from a users' group called CP/MUG, which took over the task of support by collecting and disseminating information about bugs, fixes, modifications, applications, and even the cataloging and distribution of games and other programs written or adapted by its members. Their enthusiasm and widespread expertise to a large extent made up for the absence or poor quality of the documentation and the implementation of CP/M. Without a doubt they contributed to its growth.

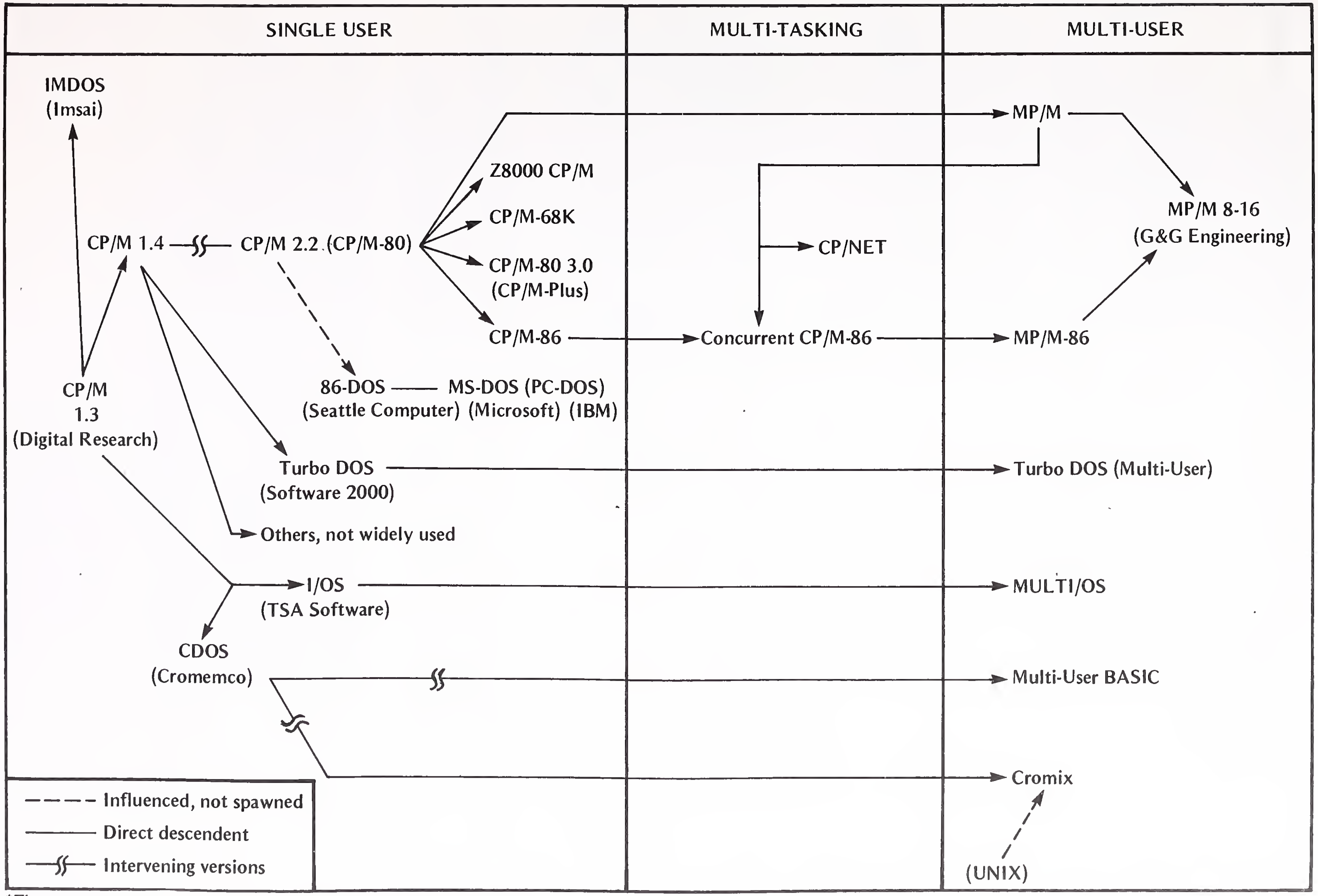
- By now CP/M has gone through many versions and spawned a number of imitators. Exhibit IV-2 shows the CP/M family tree.

2. ACCEPTANCE

- CP/M is known to be in use on more than 300,000 personal computers. If all its variants and the "bootleg" copies are counted, the number could approach a million installations.

EXHIBIT IV-2

THE CP/M FAMILY TREE*



*There may or may not be compatibility between various family members.

- One reason for this overwhelming acceptance level is the compatibility the system offers.

  - Prior to 1982 this compatibility extended only to 8080 or Z80 processors, but their number included Tandy, Heath, North Star, Cromemco, Osborne, Vector Graphic, Xerox, Wang, plus any Apple with a Softcard.

  - However, the compatibility was more in the software that managed the disk than in the disk media itself. Originally the media used the eight-inch IBM soft-sectored format, but when 5.25 inch minifloppies became popular, there was no standard format similar to the IBM eight-inch format. Goodbye, media compatibility!

  - Personal computer telecommunications had meanwhile become a popular means of software exchange, and a number of computer bulletin boards sprang up around the country. Many of these had a remote CP/M capacity to transfer files back and forth, which then could be saved on the user's own machine, no matter what the disk format.

- Now that CP/M is available for other microprocessors besides the 8080 family, source code compatibility of applications software is a lot closer to reality, although 100% object code compatibility does not exist.

  - Contributing to the compatibility advantage was the ready availability of two other early software products, as well as one later one. The early products were CBASIC (a compiled version, also marketed by Digital Research) and MBASIC (an interpreted version developed by Microsoft). The later product was the C language, for which many compilers have been developed.

- Much commercial software for personal computers and almost all personal computer systems software are now being written or rewritten in C to take advantage of the compatibility feature. Both Digital Research and Microsoft, for example, are writing the latest versions of their operating systems in C.

- In addition, other vendors have announced products that will enable CP/M (in some form) to run under UNIX (in some form). Thus users can look forward to increasing compatibility at the software level.

- As for media compatibility, the future is not so bright, but some developments will be discussed later with reference to the UCSD p-System.

- In any case, the widespread use of CP/M provided a tremendous market for the vendors of application software. There are more application software packages available to run under CP/M than for Apple, Commodore, and Tandy combined, and most of them are quite sophisticated. They run the gamut from personal and professional applications, through the range of small business functions (including vertical industry specialization), to significant information retrieval and data base management applications. CP/M has definitely been accepted.

3. IS CP/M USER FRIENDLY?

- Not very, but it is improving.

- CP/M's history, like UNIX's, began with professional programmers who were not very concerned with user friendliness. In fact, the term was not even widely recognized at the time the systems were developed.

  - There are no warnings when the user is about to commit a fatal error, such as deleting a file (or the entire disk).

- Commands are terse and not always descriptive. The file transfer function, for example, is called PIP (for Peripheral Interface Processor). It does much more than just copy files, however.

- There have always been complaints about CP/M's documentation, not only that it was inaccurate, but that it was sparse and disorganized. (The latest version (3.0) goes a long way toward eliminating that objection. It also offers an on-line "help" function.)

- Professional programmers have learned to cope with these shortcomings, but they can be mighty obstacles to inexperienced users.

• CP/M does provide a fair selection of software development aids for the assembly language programmer, and many others are available from vendors besides Digital Research. Some of these are useful to inexperienced users also. They are discussed in the following section.

4. CP/M CLONES AND RELATED PRODUCTS

• Originally CP/M was a single-user operating system, whereas UNIX was always intended for multiple users.

- These were logical choices based on the power of the processors and disks available at the time the systems were developed. Single-user systems still make sense for many situations, particularly for the 8-bit microprocessors.

- At the 16-bit level, the clock speed of the chip plus the wider data path mean that a single-tasking processor might be loafing much of the time. The economic factor is no longer that of the machine, however; it is the time value of the user, who at a minimum could be printing out one piece of material while entering something else into the machine.

- CP/M per se does not allow this, but some of its enhancements do. Two direct descendents are Concurrent CP/M and MP/M.

- Concurrent CP/M is an 8086 variant that appears quite similar to CP/M-80 except that it permits up to four "virtual consoles" for a single user. Each virtual console can be running a separate job (providing there is at least 256K, preferably more, memory on the machine), but there will be no time saving if each of the jobs is compute bound.

  - The user can switch back and forth between any of the virtual consoles at will by a single keystroke, and executing processes can pass messages to each other asynchronously. In addition, one process can initiate another.

  - Multitasking presents a threat to data integrity. Concurrent CP/M provides a "lock file" function and both a "lock record" and a "test and write record" function to help assure data integrity during updates.

- Still more powerful than Concurrent CP/M is MP/M, a multiuser, multitasking version of CP/M-86 that is upward compatible. It supports up to 16 users and provides dynamic memory allocation, queue management, and multiple printer support.

- A third alternative is CP/NET, which supports multiple users on multiple computers. One of these must run MP/M as the master node, but the others can be as small as 16K CP/M-80 sites. This is the classic star (central host) network, whose primary advantage is sharing of disk and printer resources.

- It requires an experienced programmer indeed to take advantage of any of these more powerful versions of CP/M. User friendliness is not the primary criterion for selecting one of them.

- Products have been introduced by other vendors, however, to improve the user friendliness of the basic CP/M system. Typically these products are front ends or shells, which provide a menu-oriented user interface.

  - CP+, for example, also allows the creation of a print queue and permits the user to attach a 50-character description to each file - a vast improvement over CP/M's limitation of eight characters plus a three-character, file-type extension.

  - MenuMaster allows the user to generate his own screens, including menus.

  - Power! provides a collection of CP/M utilities that access files by number rather than by typing in the error-prone command strings.

- Some effects of the cross-fertilization between products can be seen in two other front ends, known as Unica and Microshell. Both provide several useful UNIX-like features, including pipes, redirection of I/O, and file directory and search enhancements. Neither is very expensive, and each allows the programmer who is familiar (and frustrated) with CP/M to gain some of UNIX's benefits at low cost.

5. DRAWBACKS

- CP/M is cryptic and tedious to learn. But since the user must give commands directly to the operating system (with the exception described below), there is no option to learning.

  - The exception is that there is a facility for preparing a command file for tailoring turnkey applications. It is of no assistance when problems occur since there has been no error trapping facility until release 3.0, just out. Thus hundreds of thousands of users have had to deal with the dreaded message "BDOS ERR ON B: BAD SECTOR," i.e., a disk read error with no recovery options.

- The system (until the latest release) could not accommodate a change of diskettes without rebooting the operating system (warm start) because this was the only way to update the track utilization table. Uncounted hours and data values have been lost to this design defect.

- All of this would have been intolerable if there had not been such a wealth of software available for CP/M - from WordStar, SuperCalc, and dBase II to all of the popular development languages (some as subsets) and thousands of application packages. The only real difficulty may have been finding the one closest to the user's needs. Even support was widely available from dealers, user groups, and currently even a free "hotline" on The Source to provide information on updates, current versions, and a toll-free number to call for more information if needed.

- CP/M's success despite its unfriendliness to users demonstrates the relative value of application software availability. But Digital Research will not continue to depend on past successes. It has already begun to address the user friendliness issue directly as well as build bridges to the UNIX community by writing its latest products in the C language.

6. FUTURE DEVELOPMENTS

- There is a real question as to whether CP/M will be able to retain its vast installed base or whether the combined push from UNIX and Microsoft (formerly an ally, but now a strong competitor) will prove to be too much. Digital Research is certainly taking the challenge very seriously. It sponsored an industry show earlier this year devoted exclusively to CP/M. (It drew an incredible 60,000 attendees!) Several new products were shown for the first time, and the announcement of C compatibility was not underplayed.

- INPUT's judgment is that UNIX and MS-DOS (PC-DOS, really) will continue to expand the market but will not cut in much on CP/M's already installed base. The availability of packaged software is the key.

- CP/M-86 has not done well on the IBM-PC to date because of IBM's pricing differential between CP/M-86 ($200) and PC-DOS ($40), and because of its lack of applications software packages. Digital Research has now cut its price for CP/M-86 to $60, which puts it in the same ballpark as PC-DOS.

- At the same time, Digital Research has made an intense effort to make its new products friendlier, more flexible technically, and available on all the 16-bit chips.

- These measures may enable Digital Research to hold onto its present customers, but the outlook for increasing CP/M's market share is less likely. The next section describes CP/M's strongest challenger for newly installed systems, MS-DOS.

C. MS-DOS (PC-DOS)

1. HISTORY

- Sometimes it pays to plan ahead, even if you don't have a very clear idea where you may end up, only what you want to have ready when you get there.

  - That is really all Tim Paterson knew in 1978 when he decided Seattle Computer Products ought to develop a 16-bit product designed around the Intel 8086 chip. The product became the Gazelle 16-bit microcomputer. It had a unique operating system (86-DOS) patterned after CP/M but with a number of significant improvements in the way it handled disk access.

- Toward the end of 1980, IBM was searching for a company that could provide both an operating system and high-level language(s) for its planned personal computer. IBM approached Microsoft, Microsoft approached Seattle Computer Products, and Tim Paterson's 86-DOS became the heart of MS-DOS (called PC-DOS on the IBM-PC).

2. ACCEPTANCE

- It is doubtful that Microsoft would have been interested in 86-DOS if IBM had not been involved. Microsoft had been working on XENIX for the 68000, targeted at the high end of the 16-bit scale. IBM's interest, however, suggested a low-end mass market that would be hard to pass up.

- As it is, PC-DOS is the primary product, and MS-DOS, while available for other machines, is not and is not likely to be nearly as popular as PC-DOS. Differences between the two systems are minimal, but there are substantial differences between the first and second releases of the product.

- With respect to the IBM-PC, even though UCSD p-System and CP/M-86 are available, almost all purchasers choose PC-DOS rather than one of the others.

  - CP/M-86 initially promised to offer access to more software than PC-DOS did, but CP/M-86 would not run most CP/M-80 software without first converting it (through Digital Research's XLT-86 program). Extra dollars, extra time, extra complications were the conclusions most purchasers arrived at. Even CP/M-86 itself costs $200 from IBM compared to PC-DOS's $40 price tag.

  - UCSD p-System offered wider compatibility but had drawbacks of its own (which will be discussed in a later section) that made it less attractive. Also, it costs $625 (for the full development system), substantially more than PC-DOS or CP/M-86 and twice as much as IBM's own PASCAL.

- The bottom line is that MS-DOS's acceptance is really PC-DOS's acceptance, which in turn is really the explosive acceptance of the IBM-PC. By now, nearly two years after the IBM-PC was introduced, there is a great deal of application software available for it, and the original reason for wanting to choose CP/M over PC-DOS has largely disappeared.

3. IS MS-DOS USER FRIENDLY?

- More so than CP/M or UNIX, but not completely.

- The combined influence of Microsoft and IBM, and having CP/M as an unfriendly example, produced several changes that set PC-DOS apart from CP/M.

  - Documentation is much clearer. This IBM influence is now a role model others follow, but nobody does it quite as well as the master.

  - Commands are more recognizable, and error handling is greatly improved.

  - MS-DOS 2.0 provides a menu-driven user interface (that includes an on-line "help" feature) and an upwardly compatible path to XENIX. Pipes and filters are available, although they are slow. They operate sequentially through temporary disk files rather than concurrently as they do in UNIX. Files in MS-DOS 2.0 are passed to application programs in XENIX format.

  - Disk access is much faster under MS-DOS than under CP/M, and larger files can be stored on the disk. Multitrack buffers substantially reduce the number of physical disk accesses required. It is also unnecessary to reboot the operating system whenever a disk is changed (as with CP/M) because the file directories are handled differently.

- PC-DOS does not need to be configured to the particular hardware on which it will run but can read internal switch settings to determine how many disk drives, how much memory, etc., are available.

- These small features cumulatively make MS-DOS an easier product to use than the other three major systems. But several of their more powerful features have been sacrificed with the result that MS-DOS is not a very useful tool for the professional programmer.

4. DRAWBACKS

- In the user friendly race, MS-DOS is the clear winner with nontechnical computerists but not with sophisticated users or professional programmers. Even for the nontechnical user, MS-DOS is neither bulletproof nor self-explanatory. A very useful feature, such as redirecting I/O, requires patching the operating system with a debugger rather than modifying a table or furnishing a parameter.

- Advanced users will almost certainly want to take advantage of multitasking on 16-bit machines. MS-DOS does not permit multitasking, and is not likely to do so. XENIX does, but XENIX is a long step up from MS-DOS. On the other hand, Concurrent CP/M is a small step up from CP/M and is available now.

- For professional software developers, MS-DOS is not even a sensible option unless one is developing software for the MS-DOS market. The "toolbox" is simply too limited.

- Support (i.e., the lack thereof) is reported to be a significant drawback for MS-DOS users who attempt to deal directly with Microsoft. PC-DOS users can of course draw upon IBM dealer resources for support.

5. FUTURE DEVELOPMENTS

- The system of choice on the IBM-PC will continue to be PC-DOS, even with a reduced price now available on CP/M-86 from Digital Research. PC-DOS is a good product for users who do not care to invest much energy in learning the ins and outs of a more powerful system, and the number of software packages available to run under it will continue to grow.

- For those users who outgrow PC-DOS, XENIX is already waiting. Microsoft will continue to make the migration from PC-DOS to XENIX as painless as possible, and again the product is a good one.

- It would be hard to recommend MS-DOS to the user who does not wish to commit to the IBM-PC.

  - For users who do not have the IBM-PC, the CP/M family offers more powerful facilities at every step and can be made more user friendly than the native version with the addition of one of the many front ends available. The immediate advantages of CP/M are access to a broader range of application software and a more likely basis for compatibility to other systems.

  - For DEC users, UNIX or a clone (probably XENIX) might be the first choice rather than the eventual one.

D. THE UCSD p-SYSTEM

1. HISTORY

- Only one of the operating systems discussed in this report was designed from the beginning to be portable: the UCSD p-System.

- It has been implemented on the broadest range of hardware, from 8-bit to 32-bit processors. It is unique in that code produced under the p-System is portable not only at the source level but also at the object level (or more specifically, at the p-code level).

- It provides virtual memory support by automatically bringing application program code in and out of real memory when the code is needed.

- PASCAL, invented by Nicklaus Wirth, was implemented under the direction of Kenneth Bowles by the University of California at San Diego as a compiler that produced, not direct object code for the target machine, but a set of universal intermediate codes (p-code) that could either be acted upon directly (as by Western Digital's PASCAL MicroEngine) or by a p-machine emulator tailored to the target hardware. Other people have implemented other variants that may or may not be compatible with UCSD PASCAL.

  - Distribution and maintenance of the p-System was assigned in 1979 to SofTech Microsystems.

  - Originally the p-System supported only UCSD PASCAL, but now other languages are available under it as well (e.g., FORTRAN-77 and BASIC). There is no absolute link between UCSD PASCAL and the UCSD p-System since either can be used with other counterparts, under some circumstances, but obviously they were designed to work best together.

2. ACCEPTANCE

- Most users of the p-System use it with UCSD PASCAL.

- PASCAL is widely used in colleges and universities as a teaching language where speed of production generally is not a factor. Speed of production is very definitely a factor for the business user, and this more than any other feature has limited the penetration of PASCAL in the business environment.

  - Additional factors, such as pricing and available software, have also contributed to the relatively low penetration, whether under the UCSD p-System or in any other version. The fact that there are significant differences between versions does not help either.

  - This is not to imply that good software cannot be written in PASCAL. Many commercial products have in fact been implemented in PASCAL, and those who use it either like it very much or hate it. It supports the structured techniques very well, but its string and I/O handling capabilities are awkward in many variants (less so in the UCSD version, however).

- Nevertheless, many of these objections can be (and have been) overcome by clever programming design and/or restriction of language alternatives. The p-System fits on almost all 8-bit systems, for example, while the UNIX system does not.

- The slow speed may not be an objection in a single-user, single-task environment. In any case it can be speeded up by translating directly into object code rather than into p-code.

- SofTech Microsystems realizes that showdown time has come for the UCSD p-System. The 16-bit market is already slipping away to less elegant systems.

  - One factor that may help to preserve the p-System in the market is IBM's surprise selection of it as the only IBM-sanctioned operating system offered on the DisplayWriter.

- Another factor that could have an even larger impact is SofTech's Universal Medium concept, which uses software to bypass the present incompatibilities in the physical (electronic) recording of 5.25 inch diskettes. Analogous to p-codes, diskettes prepared under the Universal Medium could be read by any other computer regardless of that computer's native diskette format, as long as it also used the p-System as its operating system.

- Still another factor is IBM's decision to offer a run-time only version of the p-System on the IBM-PC for $50.

3. IS THE UCSD p-SYSTEM USER FRIENDLY?

- Not really.

- In the sense that it supports UCSD PASCAL (and the features of that language) very well, it is. But it is no easier to learn, provides little more protection from disaster, and offers less of a development toolbox than any of the other alternatives. In short, its greatest asset is its portability. This does not count for much if vendors do not choose to provide other software to run under it. So far, most applications written under the p-System are for the Apple, which offers a subset version of UCSD PASCAL on its own machine.

4. DRAWBACKS AND FUTURE DEVELOPMENTS

- Software vendors are taking a "wait and see" attitude toward the p-System. Much of their reluctance may be due to ignorance of the value of the p-System facilities. Once again elegance takes a back seat to marketing.

- If SofTech can get its marketing act together (and the recent DisplayWriter decision is a sign that it is trying), the p-System has some advantages.

    - It is very portable.

- Many computer science graduates are familiar with it, which could help recruiting efforts.

- PASCAL supports structured programming and design techniques.

- But the negatives must also be considered.

  - There is little software available for it now.

  - I/O facilities are primitive.

  - Compiling and operation are slow and, in many cases, awkward (these same complaints, however, often apply to other systems running PASCAL).

  - Debugging is miserable.

- In summary, although the p-System needs a few more years of maturing, it should be considered a viable alternative if its special characteristics fit an organization's requirements.

E. OTHER OPERATING SYSTEMS

- Two other operating systems sometimes cited as user friendly are OASIS and PICK.

- OASIS was designed to be a business operating system, not an academic or a professional programmer or hobbyist system. It offers keyed file access, print spooling, record locking, password and log-on security, an on-line "help" facility, a DBMS, and both 8-bit and 16-bit versions. For the small business user, OASIS is an attractive option. There are application software packages available for it as well as training and support.

- But even with all these advantages, INPUT does not believe OASIS will be a serious contender for the standard personal computer operating system in most Fortune 500-sized organizations. The only real reason for this is marketing. If IBM had approached Phase One Systems instead of Microsoft for the IBM-PC, we would now probably have PC-OASIS.

- The PICK operating system's future is even more of a mystery. Developed by Dick Pick in the early 1970s, it was known as Reality on the Microdata mini-computer. Several lawsuits later, both Pick and Microdata retain rights to the system. It has now been implemented on Prime, ADDS, and the IBM Series/1 computers, among others. The system features a vastly extended version of BASIC called DATA/BASIC, a query language called ENGLISH, and other powerful minicomputer-oriented features. There have been rumors of its being made available for the IBM-PC. The recent announcement of a hard disk facility for the PC makes this possibility more feasible since the PICK system requires a lot of disk space.

- INPUT believes both OASIS and PICK are "sleepers" that need to be checked from time to time. They are not front runners in the personal computer operating system race.

V IMPLICATIONS FOR IS PLANNERS

- Chapter III addressed the distinction between types of users as a major criterion in defining the term "user friendly." This chapter extends that distinction from the user to the tasks and functions to be performed. Depending upon the intended environment and application, the choice of which personal computer operating system to use can vary in importance from irrelevant to highly significant. The following sections discuss the major factors to be considered by IS planners before making a final choice.

A. BASIS FOR MIGRATION

- There are at least three strategies that may bring common personal computer operating systems under the purview of the IS planner:

  - Strategy I: Integrate personal computers into the overall IS plan.

  - Strategy II: Integrate shared logic word processing (and the more powerful standalone word processors) into the overall IS plan.

  - Strategy III: Use high-end personal computers as a less expensive route to distributed data processing.

- Each of these strategies is technically feasible now, although each has constraints that may cause a delay of a year or two.

- Strategy I is essentially a replacement for (or an extension of) an outside or in-house timesharing system whose primary user is the independent, mid-level, end-user professional.

  - This individual's computing needs include financial or statistical number crunching, graphics and text formatting, and access to data bases that may be internal or external to the corporation or both.

  - Generally, prepackaged application systems are available for much of the actual processing, but tailoring and custom program development may be necessary either for data extraction, manipulation, or presentation.

  - The quantity of data handled at one time plus its availability from an existing source (either commercial or a by-product of other in-house applications) are often overriding considerations that permit or prohibit using personal computers to perform these tasks. Network requirements are also sometimes an overriding factor.

- Strategy II, the integration of word processing into the IS plan, is a frequent approach to office automation. Its intended participants perform primarily clerical and administrative duties relating to the generation, storage, and distribution of in-house documents, comprising everything from simple notes and memos to letters and multipage reports.

  - Equipment for these tasks is usually already in place, and the overriding question is whether the equipment can be upgraded/modified/integrated as is or whether it needs to be replaced.

- The availability and suitability of local networks may also be a consideration. The ability to absorb tasks now done on terminals or telephones becomes an added plus from the standpoints of space, cost, and convenience.

• Strategy III, a low-cost alternative for implementing distributed data processing, includes as participants both data entry personnel (either as their primary duty or as an aspect of their role in sales, administration, shop floor, etc.), and also the IS staff involved in systems development or maintenance.

  - All of these people are generally performing their duties on a terminal connected to the mainframe, and the essential factor is the capability of the personal computer operating system to do or support the specific tasks required.

  - Secondary selection considerations include performance, security, and other characteristics related to the specific application.

• For all three strategies, operating system cost, support, training, and other issues are obviously important but are generally not overriding factors.

• It is not unreasonable to consider popular personal computer operating systems as components of any or all three of these migration strategies. However, since each strategy ultimately presents the possibility of personal computer operating systems interacting with a mainframe operating system, whether the personal computer operating system has this capability could become critical for the IS planner who is proposing a related corporate technical standard.

  - The intricacies of how each operating system handles this situation (if it does) and where the boundary lies between an operating system and a telecommunications system goes somewhat beyond the scope of this report.

- Suffice it to say that, at present, none of the four personal computer operating systems discussed in this report makes any provision for internetwork or interoperating system communications. But the multi-user systems (UNIX and the high-end versions of CP/M) do provide for intertask communication internally via their own networks.

- UNIX has been implemented as a task under VM/370 by Amdahl, who calls it UTS. IBM has recently announced UNIX on its mid-sized main-frames and provides a related product called CPIX on the Series/I.

- The p-System has been implemented as a task under UNIX and VAX.

- CP/M has not yet been implemented on anything in the IBM line larger than the IBM-PC, but Virtual Microsystems Inc. offers a board and program called The Bridge that make a Data General or DEC mini-computer look like a CP/M machine. This emulator concept can be (and probably has been) implemented by someone on the 370 architecture although no such product is yet commercially available.

• CP/M can also run as a task under UNIX (although not vice versa), and SofTech has announced a program that will allow the p-System to read and write files in CP/M format. Thus there are a considerable number of cross-over alternatives (both software and hardware based) already available between the major personal computer operating systems. For all intents and purposes it will soon be possible to achieve some degree of compatibility between all of them. There will obviously be cost, performance, and user hostility penalties to pay for these flexibility kludges.

MANAGEMENT PROGRAMS: Designed for clients with a continuing need for information on a range of subjects in a given area.

- Management Planning Program in Information Systems - Provides managers of large computer/communications facilities with timely and accurate information on developments that affect today's decisions and plans for the future.
- Management Planning Program for the Information Services Industry - Provides market forecasts and business information to information services companies to support planning and product decisions.
- Company Analysis and Monitoring Program for the Information Services Industry - Provides immediate access to detailed information on over 3,000 companies offering software, processing services, integrated systems, and professional services in the U.S. and Canada.
- Management Planning Program in Field Service - Provides senior field service managers in the U.S. and Europe with information and data to support their planning and operational decisions.
- On-Target Marketing - A practical, "how-to" methodology for more effective marketing problem solving and planning. Delivered to clients through workshops and/or consulting services.

MULTICLIENT STUDIES: Research shared by a group of sponsors on topics for which there is a need for in-depth, "one-time" information and analysis. A multiclient study typically has a budget of over $200,000, yet the cost to an individual client is usually less than $30,000. Recent studies specified by clients include:

- Selling Personal Computers to Large Corporations
- Improving the Productivity of Systems and Software Implementation
- User Communication Networks and Needs
- Financial Planning Systems Markets: The Next Five Years

CUSTOM STUDIES: Custom studies are sponsored by a single client on a proprietary basis and are used to answer specific questions or to address unique problems. Fees are based on the extent of the research work. Examples of recent assignments include:

- Organizing for Effective Software Development
- Corporate Plan for Utilizing CAD/CAM
- Annual ADAPSO Survey of the Computer Services Industry
- Analysis of Business Services for a Major Financial Institution
- Study of the Specialty Terminal Market
- Study of Disaster Recovery Services
- Analysis of Software Maintenance Issues
- Review of Software Product Market Opportunities
- Analysis of Network User Requirements

About INPUT

INPUT provides planning information, analysis, and recommendations to managers and executives in the information processing industries. Through market research, technology forecasting, and competitive analysis, INPUT supports client management in making informed decisions. Continuing services are provided to users and vendors of computers, communications, and office products and services.

The company carries out continuous and in-depth research. Working closely with clients on important issues, INPUT's staff members analyze and interpret the research data, then develop recommendations and innovative ideas to meet clients' needs. Clients receive reports, presentations, access to data on which analyses are based, and continuous consulting.

Many of INPUT's professional staff members have nearly 20 years' experience in their areas of specialization. Most have held senior management positions in operations, marketing, or planning. This expertise enables INPUT to supply practical solutions to complex business problems.

Formed in 1974, INPUT has become a leading international planning services firm. Clients include over 100 of the world's largest and most technically advanced companies.

OFFICES

Headquarters
1943 Landings Drive
Mountain View, CA 94043
(415) 960-3990
Telex 171407

Detroit
220 E. Huron
Suite 209
Ann Arbor, MI 48104
(313) 971 0667

New York
Park 80 Plaza West 1
Saddle Brook, NJ 07662
(201) 368 9471
Telex 134630

United Kingdom
INPUT, Ltd.
Airwork House
35 Piccadilly
London, W1V 9PB
England
01-439-8985
Telex 23116

AFFILIATES

France
La Nacelle
2, rue Campagne Première
75014 Paris
322 56.46
Telex 220064 X5533

Italy
PGP Sistema SRL
20127 Milano
Via Soperga 36
Italy
Milan 284-2850
Telex 310352

Japan
Overseas Data Service Company, Ltd
Shugetsu Building
No 12 - 7 Kita Aoyama
3-Chome Minato-ku
Tokyo 107
Japan
(03) 400-7090
Telex 26487

Sweden
Athena Konsult
P.O. Persson & Co AB
Box 22114
S-104 22 Stockholm
Sweden
08 52 07 20
Telex 17041

INPUT Planning Services for Management