DOCUMENT RESUME

ED 094 755 IR 000 955

TITLE	Northeast Academic Science Information Center (NASIC), Phase I Report (March 1973-February 1974). Volume 2.
INSTITUTION	New England Board of Higher Education, Wellesley, Mass. Northeast Academic Science Information Center.
SPONS AGENCY	National Science Foundation, Washington, D.C. Office of Science Information Services.
PUB DATE	May 74
NOTE	313p.; For related documents see IR 000 954-957
EDRS PRICE	MF-$0.75 HC-$15.00 PLUS POSTAGE
DESCRIPTORS	*Computer Oriented Programs; Cost Effectiveness; Data Bases; Demonstration Programs; *Information Centers; Information Retrieval; Information Systems; Library Reference Services; On Line Systems; Program Descriptions; *Regional Programs; *University Libraries
IDENTIFIERS	Massachusetts Institute of Technology; NASIC; *Northeast Academic Science Information Center

ABSTRACT

In the second volume of a four-volume annual report on the Northeast Academic Science Information Center (NASIC), two developmental studies are reported. The first documents an experimental, pilot operation of computer-based reference search services to users on a fee-for-service basis initiated at Massachusetts Institute of Technology as the first node of the NASIC program. The second is a study of effectiveness and cost effectiveness considerations for NASIC information services operation. The study reviewed cost factors relating to the selection of the data bases, the performance of information retrieval and dissemination systems, the choice of service centers, NASIC products and services, software considerations, communications and networking aspects, and management of information service center operation. A lengthy review of the status of on line interactive retrieval system is attached. (JY)

NORTHEAST ACADEMIC SCIENCE INFORMATION CENTER (NASIC)

PHASE I REPORT

(March 1973 - February 1974)

VOLUME 2

Submitted in lieu of the 4th Quarterly Progress Report to the Office of Science Information Service, National Science Foundation by the New England Board of Higher Education, Wellesley, Massachusetts, under Grant No. GN-37296, May 1974.

TABLE OF CONTENTS

March 29, 1974 Report ESL-R-543

NASIC AT MIT

PHASE 1 REPORT

16 JULY 1973 - 28 FEBRUARY 1974

by

Alan R. Benenfeld
Mary E. Pensyl
Richard S. Marcus
J. Francis Reintjes

The research reported in this document was performed under contract to the New England Board of Higher Education in connection with their NASIC Program funded by the National Science Foundation.

Electronic Systems Laboratory
Department of Electrical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

ABSTRACT

An experimental, pilot Operation of computer-based reference search services to users on a fee-for-service basis was initiated at M.I.T. as the first node in the development of the Northeast Academic Science Information Center (NASIC) under a New England Board of Higher Education (NEBHE) program. The development encompassed, among other tasks, selection of services, training for services, developing the initial organizational and operational policies and capabilities, publicity about available services, and the operations monitoring procedures. A fundamental philosophy is to integrate these services within the library environment where they complement traditional services. Initial experiences during a three month operational period show that (1) a demand exists for computer-based reference search services; (2) users are willing to pay, even out-of-pocket, for such services; (3) searches are often interdisciplinary and require several sources; (4) various publicity mechanisms are helpful but none so important as satisfied users telling their colleagues; (5) users like and respond positively to the in-depth, customized service and personal attention to their bibliographic needs; (6) extensive training and practice of Information Specialists is necessary to attain desired levels of service quality; (7) integration of these services within [illegible]y environment may require organizational and staffing accommoda[illegible]dition to the personal cooperation, good will, and enthusiasm of participants.

ACKNOWLEDGEMENT

This report represents the first phase of effort by a NASIC at M.I.T. project team. Two vital components of the team are the M.I.T. Libraries and the Information Processing Services. Their respective directors, Natalie N. Nicholson and Robert Scott, take active interest in the development of NASIC on the M.I.T. campus and the contributions by them and their staff are valued.

Special acknowledgement is given to our Information Specialists--Marge Chryssostomidis, Pat Gordon, Ann Longfellow, Jackie Stymfal, Nancy Vaupel, and Susan Woodford--who provide the service that NASIC is all about, whose enthusiasm and committment are infectious, and to whom the rest of the team owes a special debt of gratitude.

Our appreciation is also extended to Phillip Piper, the Assistant to the NASIC Coordinator, for his help in the coordination of the various activities supporting the service.

Progress has been helped by communications among team members. One major aspect of the communications process is the memoranda among participants. Appreciation is extended to the project secretaries, Susan Garland and Karen Parry, for their expeditious handling of the documentation.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION AND OVERVIEW

An experimental, pilot operation of computer-based reference search services to users on a fee-for-service basis was initiated at M.I.T. on 15 November 1973. It marks a major milestone in the development of the Northeast Academic Science Information Center (NASIC). NASIC development is supported by a grant from the National Science Foundation to the New England Board of Higher Education (NEBHE). Development of a pilot operation at M.I.T. is supported by subcontract from NEBHE to M.I.T. The NASIC at M.I.T. project team includes staff from the M.I.T. Libraries, the Electronic Systems Laboratory, and the Information Processing Services. This report covers the work performed on NASIC at M.I.T. from 16 July 1973 through 28 February 1974. This period falls within Phase 1 of NASIC.

Services for a fee have been provided to more than 60 users. Although the number of users is small at this still early stage, the ranks of users continue to grow with the increasing publicity about the availability of NASIC services. The effort to reach this stage by the M.I.T. project team encompasses, among other tasks, selection of services, training for services, developing the initial organizational and operational policies and capabilities, the publicity about available services, and the operations monitoring procedures. A philosophy fundamental to this effort is to integrate such services within the library environment. Details of the work accomplished on each task are contained in subsequent sections. The reader is referred in particular to Task 12, Monitoring and Analysis of Service Operations, for an extended analysis of results.

It is worthwhile highlighting the more important findings to date of this development and testing effort. They are:

1. A demand exists for computer-based reference search services available on a fee-for-service basis.

2. A measure of the strength of the demand is that many users are willing to pay out-of-pocket for such services, although the majority of users to date have access to contract or grant monies.

3. Mechanisms need to be established to support large-scale use of NASIC services by undergraduates and others who do not have access to grant or contract monies.

4. These services complement but do not replace more traditional search modes.

5. A significant number of user search problems are interdisciplinary in nature and may require searches in more than one data base.

6. Various publicity mechanisms are successful but, not unexpectedly, the most important one seems to be satisfied users telling their colleagues about NASIC.

7. Users like and respond positively to the in-depth, customized service and personal attention to their bibliographic needs.

8. To provide quality service Information Specialists need to receive training including fairly extensive practice searching. It takes additional experience for an Information Specialist to become fully confident, adept, and at ease with his or her professional ability to provide an intensive customized computer search.

9. To date, a typical NASIC user appointment lasts 70 minutes. Somewhat more than half that time, 37 minutes, is spent in on-line connection to the computer. The average printout request results in 39 pages containing 131 citations. The total cost of a typical search is $50.47 composed of $34.90 for computer connection and search and administrative charges, $ 9.36 for the time of an Information Specialist, and $ 6.21 for off-line printout. However, only 60% of users request offline printout.

10. Integration of these services within the library environment may require organizational and staffing accommodation in addition to the personal cooperation, good will, and enthusiasm of participants.

The discussion by task which forms the body of this report relates to the M.I.T. environment. The NASIC at M.I.T. organization reflects this environment as well as the dual purposes of providing user services and of developing and testing methods and types of services. As such, the current setup at M.I.T. represents only one of many organizational models; we want to experiment with other models. The particular method that other institutions use to organize NASIC computer-based reference services on their own campus must reflect their own local environment and needs. The M.I.T. experience ought to be of help to other universities in setting their own course.

There are several aspects of the work accomplished within the short span from commencement of M.I.T. effort on 16 July 1973 to 28 February 1974 which cannot be adequately expressed either in the highlights above or in the following summary by task. NASIC is an addition to library services at M.I.T. and not a replacement for traditional services. This is expected to hold true for NASIC service sites at other educational institutions in the region. The extension of services in a short time span to include NASIC has caused an additional burden to be carried by the M.I.T. Libraries. The M.I.T. Libraries did not have benefit of time for budgetary planning. Nevertheless, whenever problems arose, the long view of the situation was kept in mind by all concerned. Some of these problems may also arise at other sites and NASIC and the participating institutions need to give it full consideration. Regular work loads and personnel assignments in the M.I.T. divisional libraries have been disrupted during the transition period to build, operate, and continue further development and growth of NASIC services. The Information Specialists themselves have shouldered most of the burden, but in the process, and with their enthusiasm, they have shown their mettle. The administration and staff of the M.I.T. Libraries have enthusiastically supported efforts to make NASIC services a success while tolerating the transitory hardships to personnel and regular work loads associated with the magnitude of their effort.

Despite the usual trials in this development effort, it is obvious to us at M.I.T. that no matter how much or how little is available in dollars or in time, people still make the difference. Enthusiasm and committment are prerequisite for any university library staff that wishes to extend its services to computer-based reference and retrieval.

At M.I.T., we have wrestled with an important question. If each university sets its own operational course what then is the role of the Northeast Academic Science Information Center? NASIC does have a role to play, a very positive role. We have arrived at this answer by careful review of the activities and events to launch and carry forward NASIC services at M.I.T. in which the M.I.T. departments that have cooperated in this venture can be likened to a NASIC network in miniature. The major functions of a strong central regional NASIC organization are:

1. Advise academic institutions on preparing for, implementing, and publicizing computer-based reference services.

2. Offer programs to train staff to levels of competency in understanding and providing such services that extend beyond current programs of retrieval system suppliers.

3. Provide a central capability to search those systems or data bases that are only of infrequent use to an academic institution.

4. Provide a strong, collective voice for the region in dealing with retrieval systems, data base suppliers, terminal manufacturers, or other external agencies.

5. Provide a mechanism for disseminating within the region updated information and solutions to problems of common interest.

In short, a regional NASIC is needed to function as a strong user association, a center with the expertise and time to daily make suggestions and provide feedback among individual academic institutions and a variety of diverse information or equipment suppliers.

Our prognosis for NASIC remains optimistic. It is possible for an organization to implement these services entirely on its own; but in so doing, more of its resources will be required in order to fully realize the benefits from extending its services to both current and new library users. These services are exciting because they ultimately touch upon, indeed should be integrated with, a wide spectrum of information services, but they are also exacting in their implementation if their potential is to be realized. A NASIC that functions as a strong central association of members could considerably ease this process with consultation, with training, with back-up services, with collective voice to suppliers, and with feed-back to members.

II. DESCRIPTION OF PROJECT WORK BY TASK

Obtaining Service Data From External Suppliers (Task 1)

Data describing the characteristics, modes of access and costs of available external online and offline bibliographic services was gathered and reviewed. This data gave assistance to the selection of retrieval systems and data bases for initial NASIC services. Sufficient data for both on-line and off-line modes is on-hand to assist in the selection of additional data bases as sources of NASIC services. However, the external retrieval systems and data bases undergo continual modification and change so the data gathering effort also continues in order to keep abreast of such changes.

Part of the data gathering effort has taken place in conjunction with the NEBHE-ARL site survey study. M.I.T. has participated directly in site surveys at the University of Georgia, Illinois Institute of Technology Research Institute, University of California at Los Angeles, Ohio State University, University of Florida, and the North Carolina Science and Technology Research Center. M.I.T.'s role in these visits has concentrated mainly on the current and planned services and operations of these centers.

A visit with the staff of the System Development Corporation yielded new information and data on ORBIT. M.I.T. has received visits from representatives of SDC, Lockheed, and the University of Georgia which have also been beneficial in obtaining recent data.

Selection of Services (Task 2)

The ultimate long-range NASIC goal is to provide access to most computer-based information services. Four areas of interest have been considered: (1) data bases, or machine-readable bibliographic or other surrogate records; (2) data banks, or machine-readable numerical data, either raw or reduced; (3) text files other than surrogate records that are computer-stored; (4) non-computer stored (traditional) information.

The first area, data base or bibliographic access, was a given for NASIC services and it forms the core of NASIC activity for the forseeable future. NASIC will use the services of existing retrieval systems and data bases rather than develop processing capabilities of its own. While access to all such bibliographic services is desirable and necessary for comprehensive coverage, an initial set of services had to be selected. Other bibliographic retrieval services will be phased in over time. Time to accomplish training of personnel is a critical element in the introduction of all such services.

NASIC services at M.I.T. were initiated online with the SDC ORBIT system and offline with the University of Georgia system. The SDC ORBIT system was selected for several reasons. It is somewhat easier to learn to use the ORBIT retrieval system than most others. It is accessible via the TYMNET network making communications easier. It had, last August, more data bases of interest to the M.I.T. community. It is essentially the same retrieval system as the National Library of Medicine MEDLINE system access to which was being arranged concurrently but independently by the M.I.T. Science Library. The University of Georgia system was selected because it has the largest array of data bases in the country, many of which are searchable in both retrospective and current awareness modes.

Specific data bases available on those systems were also selected for initial operations. The on-line data bases include the Chemical Abstracts base (CHEMCON), plus data bases in education, linguistics and information science (ERIC) and in business, management, and economics (INFORM).

These data bases are of interest to large segments of the M.I.T. community. Other data bases available at that time on ORBIT but not selected were in medicine (MEDLINE) and agriculture (CAIN). There is a large interdisciplinary interest at M.I.T. in the medical data base but SDC's service was not selected because concurrently with NASIC services, MEDLINE service is being made available (and at lower cost) by the M.I.T. Libraries through arrangement with the National Library of Medicine. However, a cooperative spirit and effort exists at M.I.T. in providing NASIC and MEDLINE services to the ultimate benefit of both the user community and the growth in usage of computer based retrieval systems in general.

The offline data bases initially selected for access from the University of Georgia are the Chemical Abstracts Condensates and ERIC. This selection complements or extends the data base time period coverage and type of service provided by the on-line data base. Thus, a full range of services is being provided for each data base whenever possible. Retrospective searches are available both on-line and off-line. Current awareness searches are available off-line, although ORBIT has plans to provide such a capability. The relatively new INFORM data base is the only SDC data base that does not have an off-line counterpart at Georgia.

Since the time of selection of initial data bases for NASIC, SDC nas made available data bases in engineering (COMPENDEX), geology (GEO-REF), and is expected soon to make available government reports (NTIS), and a citations index (ISI). NTIS, CAIN, COMPENDEX and GEO-REF are currently available off-line at Georgia. While a detailed plan for phasing-in other data bases and retrieval systems is in preparation, these particular data bases are of interest to the M.I.T. community and there is a high probability of adding them next. A contributing factor in the choice is that specialist training in the use of these data bases can be accomplished sooner than would be the case if, in addition, a new host retrieval system first had to be mastered. However, the effort in testing additional retrieval systems for eventual use is continuing and is now being aided by the use of real search problems. Indeed, comprehensiveness of bibliographic data base coverage is one of the most valuable features that NASIC can provide.

The NASIC effort is being concentrated on bibliographic data base access. NASIC access to data banks of machine-readable numerical data, either raw or reduced, such as, for example, the census, was considered. However, numerical data bank access was deferred for the present in order not to dilute the available resources necessary for successful training in data base access and because information specialist personnel with backgrounds additional to that of data base specialists, particularly in programming, are required. However, M.I.T. and NEBHE studies of this area do indicate the desirability of having effective interaction and referral between data bank specialists and data base specialists.

Further consideration of the third area, machine-stored, full-document text files has also been deferred because of the very limited number of sources at the present time.

It is particularly important to conclude this summary on selection of services with a note that a conscious and continuing effort is being made to develop an effective interface between the NASIC computer-based information services and both traditional searching and document delivery systems. While NASIC is primarily concerned with computer-stored information, non-computer-stored information is of particular concern for at least two important reasons, both of which are highly likely to influence a user's opinion of the effectiveness of NASIC services. First, most, if not all, machine-stored data currently available, have limited retrospective capabilities (on the order of a few years for most bibliographic sources). Some number of users will have need to search further back in time. Hence an effective interface needs to be designed between computer-based information services and traditional searching. Second, both traditional and machine-based bibliographic searches on document surrogates usually generates a need to obtain access to the full documents. Hence, an effective bridge between machine-based NASIC services and a document delivery system (from holdings determinations through delivery) also is highly desireable. The initial NASIC effort is giving full consideration to the interface with traditional resources. For example, the reference staff is being given sufficient information to be able to refer users to NASIC; the reverse situation also arises because some users who come to

NASIC have information needs answered in part by, or only by, traditional sources. In the next phase, we plan to experiment with document delivery services as a means for following through and completing the retrieval function.

The ultimate goal is, of course, an integrated set of information services available to a user community at M.I.T. or at any other NASIC site.

Library Staffing and the Selection of Information Specialists (Task 3)

An initial core of five Information Specialists were selected, one each from the reference staffs of the five Divisional Libraries at M.I.T., to provide NASIC services part-time. A sixth Specialist was subsequently brought on board. The Information Specialists are:

Ms. Marge Chryssostomidis	-	Barker Engineering Library
Mrs. Pat Gordon	-	Science Library
Ms. Ann Longfellow	-	Rotch Library
Mrs. Jackie Stymfal	-	Dewey Library
Ms. Nancy Vaupel	-	Humanities Library
Ms. Susan Woodford	-	Science Library

The selection of Information Specialists reflects the general organization of the M.I.T. Libraries which are decentralized into five divisions corresponding to each of the five schools (Engineering, Science, Architecture and Planning, Management, Humanities and Social Science) at M.I.T. Each Divisional Library could be expected to interface in some way with NASIC. To handle the interface and to generate and maintain interest in NASIC throughout the total Library system, the Library Administration selected a part-time Information Specialist from each Divisional Library rather than one or two Specialists for the total library system.

The Library Administration asked the Head of each Divisional Library to recommend one of their staff members for the job. Although no formal selection criteria were established, in point of fact the most important informal selection criteria were previous experience with computer-based services and/or a high level of interest or enthusiasm in such services. Selection of Specialists was accomplished without any difficulties. Two of the initial core of five Specialists had previous experience with computer-based bibliographic services, in particular, Intrex, MEDLINE, and several batch systems. Personal interest and enthusiasm from the other Specialists did indeed rank high, (one concurrently took a programming course at M.I.T.).

The M.I.T. Libraries are providing, in parallel with NASIC services, access to the National Library of Medicine MEDLINE system. The Libraries sent one staff member from the Science Library to participate in the 3-week Medline training program. While NASIC and MEDLINE services are parallel services, from a user's viewpoint, this represents proliferation and is a potential source of confusion for him. It seemed highly desirable to provide a single point of access to any computer-based information source for the M.I.T. user community. A coordinated effort can be and was implemented with greater convenience to both user and staff while also providing appropriate sponsorship credits and maintaining appropriate cost allocations for contractual obligations. To this end, the initial core of five NASIC Specialists was enlarge such that the MEDLINE trainee was additionally trained to provide NASIC services also. MEDLINE is a member of the ORBIT retrieval system family and this fact has facilitated the coordination.

Although an initial core of NASIC Information Specialists had to be selected, we would expect that, in time and with increased user demand, the entire reference staff of the M.I.T. Libraries would be trained to provide computer-based reference and retrieval services. Indeed, several members of the reference staff have expressed such an interest.

Development, implementation, and the management of operational activities within the Libraries requires coordination, a fact recognized from the outset in pre-contract discussions. The decision later to use a decentralized staff of Specialists underscores the role of a Coordinator. Ms. Mary Pensyl is the NASIC Coordinator and she has had previous experience in the use and promotion of computer based information services. The NASIC Coordinator reports directly to the Library Administration. As we move into Phase II, the NASIC Coordinator will be less concerned with development and implementation but instead, address more attention to publicity and marketing.

Subsequent development of service operations at M.I.T. pointed up a distinct need for support capability. An Assistant to the NASIC Coordinator, Mr. Phillip Piper, was hired to carry out the duties of manning a telephone to receive and answer inquiries, to schedule and carry out all arrangements for service appointments, to maintain user files, to handle billing and accounting procedures that flow through the Coordinator's Office, to distribute all off-line printouts sent to this central office, to gather data on the service operations and to assist in its reduction.

Access to External Services (Task 4)

Accounts have been obtained by NEBHE for use by M.I.T. to access the SDC ORBIT system in an operational service environment. Other SDC accounts obtained by M.I.T. are being used for training and experimental testing purposes. Two accounts have been obtained by NEBHE for use by M.I.T. to use the University of Georgia System; one account will be used for service operations and the other for training and experimentation. Accounts have also been obtained by M.I.T. to access the Lockheed DIALOG and the Battelle BASIS-70 systems for testing and training pruposes.

Training Program (Tasks 5 and 8)

A training program began in late August 1973 and continued through 15 November 1973 in order to achieve by the time of the initiation of NASIC service operations a comfortable level of understanding and ease of use of each of the initial NASIC retrieval systems and data bases by all of the Information Specialists. Training was particularly intensive in the few weeks immediately prior to initiating services. (At M.I.T., MEDLINE is being coordinated with NASIC as a service activity, but MEDLINE was not part of the NASIC training program. MEDLINE services are currently being provided only by the two Specialists from the M.I.T. Science Library both of whom had previously received MEDLINE training at NLM. Their MEDLINE service activities are in addition to their NASIC service activities.) Since 15 November 1973, a small amount of continued training has occurred to keep abreast of changes in the initial retrieval systems and data bases, and to discuss and review the operational experiences of the Specialists. Of course, additional retrieval systems and/or data bases that are to be accessed as a NASIC service will require additional training effort equivalent to comparable segments of the initial program.

The initial training program required approximately 130 man-hours (3.7 man-weeks) of effort per Information Specialist. About 40 man-hours effort per Specialist were devoted to the first module of the training program. Bibliographic and information science concepts independent of specific retrieval systems and data bases were covered and a general philosophy of service was discussed. Major topics in Module A were:

-- search problem elicitation

-- profile or search concept development

-- Boolean concepts

-- natural language and controlled language characteristics

-- retrieval effectiveness concepts

-- other search strategy techniques

-- user search satisfaction criteria

The second and third modules together represent the bulk of the effort and together required 90 man-hours of effort per Specialist. The second module covered the specific protocols, commands and other characteristics of the ORBIT and the University of Georgia retrieval systems, and the specific characteristics of the Chemical Abstracts, ERIC, and INFORM data bases as they are applied on those systems. The second module also included hands-on experience at terminals for ORBIT (approximately 10 hours of connect time per Specialist) and the development and running of profiles for Georgia. Several real users participated in the training effort providing us with real search problems and the opportunity to conduct reference interviews and to obtain feedback on terminology and search strategies. The third and smallest module was concerned with the NASIC at M.I.T. pilot service operational procedures and covered administrative, accounting, service, feedback and analysis, and similar matters.

The training program was an amalgam of lectures and discussions, individual study, practice online sessions in pairs and individually, user interviews, and offline profile development. Training materials included assigned readings, system and data base descriptions and manuals, system newsletters, and retrieval aids such as thesauri. NEBHE personnel frequently attended the lecture and discussion portions of the program. In addition a day and a half long workshop was held at M.I.T. by Mrs. Margaret Caughman of the University of Georgia to review their system and the two data bases being accessed, to cover profile weighting, and to review profiles submitted by M.I.T. for training purposes. Mr. Ran Hock of the University of Pennsylvania also attended the workshop.

The topics covered during the initial training program are to be reviewed during the next phase of NASIC and reassembled for use in a program to train Information Specialists at other NASIC sites to a similar level of preparation and understanding in the use of computer-based information services. It is expected that the training program itself will be recast to better meet the needs of geographically dispersed trainees. It is anticipated that greater use will be made of guidelines, of "tutoring" by experienced Specialists, of apprentice search sessions. A lot of information must be covered together with a lot of practice, but it is anticipated that the training program can be so segmented as to allow time for a prospective Information Specialist to absorb the material.

Development of Operational Procedures at M.I.T. (Task 6)

An intensive effort to develop and set up operations for a pilot service took place along several fronts. The efforts can be categorized as: (a) library and other sites for service; (b) terminals for on-line services; (c) service modes; (d) service charges; (e) accounting and billing mechanisms; (f) data gathering for management information and statistical purposes. These efforts on initial development have essentially been concluded Details are given on the following pages. Modifications to the initial procedures will be made on a continuing basis as further experience, feedback and analysis may dictate.

Library and Other Service Sites (Task 6A)

NASIC services are being offered through each of the five divisional libraries at M.I.T. Each library except the Science Library was surveyed together with the Head of that library in order to select particular locations within the library suitable as a NASIC service site. Criteria used to select a site included: visibility of on-going NASIC services to other library users; sufficient physical space for both users and for small group demonstrations; electrical and telephone installations; noise level, general environment, and traffic flow; ease of access to the reference collection; ease of access to quarters for secure storage of terminals. As you might expect, some tradeoffs had to be made. In some instances, particularly where large expenditures for physical improvements might have been required for an otherwise desireable location, temporary sites are being used instead pending review based on operational use. The site being used in the Science Library for NASIC services had been previously selected for MEDLINE service and it meets all of the criteria established for a NASIC site. That site has been strikingly and attractively decorated by the M.I.T. Libraries and it serves as a model for the other libraries. Sites outside the Libraries at which NASIC services are to be offered upon user request and on a test basis include the on-campus offices and laboratories of the user community. The object is to increase further the convenience of access to NASIC service for users.

Terminals for On-Line Services (Task 6B)

Texas Instruments Silent 700 thermal terminals with upper/lower case capability were selected for initial NASIC services. They are quiet, reliable, portable, and operate at 30 characters per second output, all characteristics that make them highly suitable for operation within a library environment. Two such terminals have been leased by the Project and, together with a third such terminal leased by the M.I.T. Libraries, are being shared among the five Divisional Libraires for NASIC services. A fourth terminal leased by the Electronic Systems Laboratory is being made available to NASIC for back-up purposes. While these are portable terminals, they do weigh-in at 30 pounds. Terminal logistics have been a bit of a nuisance. On occasion, the Information Specialist and user have travelled to the terminal rather than the reverse. Until user demand justifies additional terminals such that there will be at least one per library, we may, in the next phase, reduce the number of locations at which NASIC services are available. In addition, Specialists seeking online practice training with additional data bases have occasionally been hampered because the shared terminals were not close-by.

Service Modes (Task 6C)

The reader may appreciate that in preparing for NASIC services, we had to work with two unknown quantities, namely, initial user demand and its rate of growth. Therefore, considerable attention was given to determining the best modes of pilot service operation that would simultaneously permit convenience of service to users; controlled growth of service operations to meet demand when and where it arises; control over services for monitoring and analysis purposes; flexibility for easily modifying service procedures; flexibility in working with time slots when only certain on-line data bases are available; flexibility for both NASIC and for the Information Specialists in interfacing with traditional library activities; and flexibility in adding new operational services.

An appointment basis for service was selected over on-demand methods as it provides the best accommodation of both flexibility and control, particularly with limited personnel resources. In addition, the appointment places NASIC services on a more professional footing because the user is assured of undivided attention and service at the appointed hour--not a small matter when he is paying for service.

Each of the Information Specialists has been trained to work with each of the initial NASIC retrieval systems and data bases (MEDLINE excepted). A user may obtain service at any Divisional Library, or upon request, at his office or laboratory. A logistics schedule for appointments was drawn up based upon personnel and the time of day that online data bases are available. A Specialist, although attached to one library, may be asked to meet an appointment elsewhere. Appointments are scheduled centrally through the NASIC Coordination Office.

The initial appointments schedule (see Figure B1) theoretically is able to accommodate a total of 75 hours of service, 54 hours on-line and 21 hours off-line. The schedule allows the user some flexibility in arranging an appointment. The total is twice the number of hours we anticipated as being actu lly necessary to service a demand rate anticipated to grow to 25 users per week after the first few months of operation. Each Specialist has been scheduled to cover at some point each of the data bases and systems (on-line and off-line, and retrospective searches as well as current awareness profiles) although some weighting has been done to provide more hours of service for the chemistry data base by the Specialists from the Engineering and Science Libraries. MEDLINE services have been integrated into the appointments schedule but such service is available only from the Science Library Specialists trained at NLM. All appointments are nominally scheduled for one-hour but the schedule allows an additional half-hour to catch latecomers and run-over sessions, as well as to allow some time for the Specialist to complete a write-up about the appointment for later monitoring and analysis. Appointments are scheduled for the time slot and location most convenient to the user. With a moderate demand level and current staff committments, we anticipated no more than a two or three day wait. The appointment mode has proven to be useful in practice but detailed discussion of the match between the plans on which service appointments are based and our real operational experience appears below in Task 12, Monitoring and Analysis of Service Operations.

In Phase 2, we want to experiment with other modes of service. These include delegated searches, SDI online services, on-demand searches without an appointment, user self-searches, service in an office or laboratory, and document delivery services. Document location and delivery services would provide an important follow-through to any of the search modes since the search result only provides references to documents. Document delivery also promotes integration of computer-based searches with more traditional library activities.

Service Charges (Task 6D)

A pricing structure for NASIC services at M.I.T. was developed. The structure is a compromise between the complexities associated with accurately predicting all costs when we have no historical cost data to go by, and a toc simple flat rate that would reflect real costs inaccurately. The pricing structure has four main components whose sum is the total cost to the user. The components are (1) a direct computer search cost to which a surcharge is added for recovery of administrative costs, (2) a direct charge for the time of the Information Specialist, (3) a direct charge for the cost of off-line printouts, (4) charges for special services.

(1) Computer search costs are established as rates based upon either the number of hours of terminal connect time for on-line services, or on the number of years of data base coverage searched for off-line services. The rates are retrieval system and data base dependent. Table 1, a draft price list, shows the rates in effect from 15 November 1973 to 31 January 1974. Table 2 gives the rates in effect for February 1974. The change in the rates for some data bases reflects changes in the supplier's rates. The computer search cost rate includes (a) the actual rate charged by the external service less any discount provided to NASIC; (b) a rate surcharge added by NASIC at NEBHE to support the regional organization, this surcharge initially equal to the discount received from the external service; (c) a rate surcharge added by M.I.T. to defray its expenses in providing central personnel, telephones, terminals, and materials necessary for service. The M.I.T. administrative surcharge included in the search cost rate is currently set at $12 per connect hour or $6 per profile-year. The M.I.T. administrative surcharge of $12 per connect hour was derived by assuming adhievement within the first year of an operational level of 125 users per month, with the average user requiring an hour of Specialist time and 30 minutes of terminal connect time. It was also assumed that this volume would require a Coordinator at 20% full-time, and an Assistant to the Coordinator at 60% full-time. Monthly costs were estimated for salary and benefits for the Coordinator and the Assistant,

Table 1

INITIAL RATES EFFECTIVE 15 NOVEMBER 1973-31 JANUARY 1974

M.I.T./NASIC PRICE LIST (DRAFT 11/29)

The following prices for services provided at M.I.T. for NASIC and other information services will be in effect until February 28, 1974.

A. SPECIFIC PRICES

1. Information Specialist: $8/hr. (minimum charge $5) [This charge is currently being credited -- see (B) below]
2. Offline Printouts: $0.10/page (output onto 4 x 6 cards: $0.05 per card)
3. Special Services: (prices not yet worked out)
4. Computer Search: (see Table below -- charges in dollars) [minimum charge for computer search: $5]

Data Base	Type of Computer Search		
	On-line*** (per connect-hour at terminal)	Current Awareness*** (annual subscription)	Offline Retrospective**** (per year of data base searched)
CA-Condensates*	$55	$370	$166
ERIC*	$44	$ 86	$ 76
INFORM	$67	OL	OL
MEDLINE**	$18	OL	OL

OL = service available only online

*The chemical and education data bases are each divided into two parts for offline searching. If your problem can be handled by only one part, pricing can be cut in half. See brochures on these data bases for details.

**The MEDLINE (Medical Online) service is being provided by M.I.T. in cooperation with the National Library of Medicine which subsidizes the major portion of the costs. One third of the MEDLINE computer search charge ($6/hr.) goes to NLM, the remainder goes to M.I.T. NASIC does not participate in MEDLINE service and collects no payment for it.

***Charge prorated if less than hour or year of service.

****Offline retrospective searching for the current year is charged on the basis of the current awareness rate -- pro rata for that portion of any incomplete volumes to be searched.

B. INTRODUCTORY CREDIT OFFER

As an introductory offer M.I.T. is crediting each new user's account with a total credit of $50 which can be used to defray the charges for the information specialist. This credit is limited to M.I.T. users and must be used before the end of this academic year in June, 1974.

Table 2

SERVICE RATES EFFECTIVE FOR FEBRUARY 1974

NASIC* AT MIT

AUTOMATED BIBLIOGRAPHIC SERVICES FOR RESEARCH

COST OF SERVICES

Effective February 1, 1974, costs for NASIC search services are as listed below.

Information Specialist: $8/hr (minimum $5). This charge is currently being credited for MIT faculty, students and staff -- see (A) on reverse side.

Computer Search: As shown in the table below, there are two rates, one for educational and government users (EDUC) and one for commercial users (COMML). In either case the minimum charge is $5.

Off-line Printouts associated with the computer search may involve an additional charge as shown in the table.

	ON-LINE SEARCH per terminal connect hour		Off-line Printout
DATA BASE	EDUC	COMML	
CA-Cond.	$ 67	$ 82	$.08/cit.
ERIC	$ 47	$ 62	$.08/cit.
INFORM	$ 67	$ 82	$.10/cit.
MEDLINE(B)	$ 18	$ 18	$.10/page

	OFF-LINE RETROSPECTIVE per year searched		Printout
DATA BASE	EDUC	COMML	
CA-Cond.			citation only
odd or even	$ 86	$ 96	paper - free
odd and even	$166	$176	card - $.02/cit.
ERIC			abstract & cit.
RIE or CIJE	$ 41	$ 51	paper - $.10/cit.
RIE and CIJE	$ 86	$ 96	card - $.12/cit.

	CURRENT AWARENESS (C) annual subscription		Printout
DATA BASE	EDUC	COMML	
CA-Cond.			citation only
odd or even	$190	$200	paper - free
odd and even	$370	$380	card - $.02/cit.
ERIC			abstract & cit.
RIE or CIJE	$ 46	$ 56	paper - $.10/cit.
RIE and CIJE	$ 86	$ 96	card - $.12/cit.

For further information see the detailed NASIC at MIT brochures or contact the NASIC Coordinator's office:

253-7746

Room 10-400

(OVER)

SAMPLE COST CALCULATION

Assume an INFORM search by an academic user from MIT takes 45 minutes of the time of an Information Specialist (3/4 x $8 - $6.00), 30 minutes of terminal connect time (1/2 x $67 = $33.50) and results in 42 citations being printed off-line (42 x $.10 = $4.20). Total cost is $43.50, of which the $6.00 for the Information Specialist's time will be credited, leaving a net cost of $37.70.

A commercial user's cost for the same service would be $6.00 + (1/2 x $82) + $4.20 = $51.20.

EXPLANATORY NOTES:

(A) During the 1973-74 academic year, each account of a user from MIT will be credited with $50 applicable to charges for the time of the information specialist.

(B) The MEDLINE (Medical Online) service is being provided by MIT in cooperation with the National Library of Medicine which subsidizes the major portion of the costs. One third of the MEDLINE computer search charge ($6/hr) goes to NLM, the remainder goes to MIT. NASIC does not participate in MEDLINE service and collects no payment for it.

(C) The charge for Current Awareness is prorated for subscriptions of less than one year.

For further information, see the detailed NASIC at MIT brochures or contact the NASIC Coordinator's office:

25[illegible]

Room 10-400

*NORTHEAST ACADEMIC SCIENCE INFORMATION CENTER

A Program of the New England Board of Higher Education. NASIC is supported by the National Science Foundation under Grant No. GN37296.

2/74 - 4M

(FRONT) (BACK)

for terminal rental, for terminal paper, for telephone and message units, and for miscellaneous supplies. These costs were pro-rated among the assumed monthly user volume and translated into a connect-hour charge on the basis of the assumed half-hour average connect time per user. The M.I.T. administrative charge of $6 per profile year for offline searches was derived in the same way except that terminal cost and terminal supplies were excluded, and it was assumed that the average user would obtain a one-year current awareness profile or a one-year retrospective search. Beginning 1 February, NASIC services were extended to users affiliated with industrial or commercial orgainizations. At the request of NEBHE, a surcharge to this group of users was instituted of $15 per connect hour for online services or $10 per profile year for offline services. The industrial user surcharge is passed completely to NEBHE to help defray the real costs in developing NASIC. The National Science Foundation grant is paying the development costs for the academic community. There is no industrial surcharge on MEDLINE service because it is not a part of NASIC. The total charge to a user for the computer search cost component, including all applicable surcharges, is pro-rated for online searches under an hour, or for offline searches for less than a year, but there is a minimum charge of $5.

(2) Information Specialist time is charged at the rate of $8 per hour during an appointment with a minimum charge of $5. No charge is made when the Specialist and user together decide that a NASIC service would be inappropriate for the user's problem, or when a potential user seeks general information about NASIC services. The hourly rate for the Information Specialist was derived assuming a one-hour user appointment and an additional 10 minutes preparing for user sessions or for post-session clean-up. The rate is based on salary plus employee benefits.

(3) Off-line printout charges for the period 15 November 1973 through 31 January 1974 are shown in Table 1 and for February 1974 in Table 2. The initial rates were all on a per-page charge and, for simplification, had been set at a uniform cost for all data bases. Beginning 1 February 1974 new rates were set to reflect a significant change in the actual charge basis used by SDC. Their rates are now currently on a per-citation basis. At the

time the SDC change was impending, we thoroughly reviewed the basis of the initial NASIC printout charge in terms of cost recovery. We concluded that the printout component of the NASIC price structure for a given data base should be in concert with the structure of the data base supplier. Thus, the printout component is now retrieval system and data base dependent and is similar to the computer search component except that the printout component contains no surcharge for administrative costs. Recovery of overhead associated with printouts is to be included later with other indirect costs in the connect-hour surcharge.

(4) Special services will be charged for in addition to the above costs. Such services may include, for example, document reproduction and delivery. No special services have been developed during the Phase 1 period nor have costs and prices for anticipated services been investigated.

In order to aid in the introduction of NASIC services to the M.I.T. community, the M.I.T. Libraries are forgiving users the charge for the Information Specialist time up to a maximum of $50. This is the equivalent of about six hours of appointments. The credit is limited to M.I.T. users and must be used before the end of the academic year in June 1974.

Some readers may be interested to know that in accordance with a new National Library of Medicine policy to allow all MEDLINE users to obtain a standard mode of service at a standard price, we have announced the availability of a "standard" MEDLINE search at a fixed fee of $7.50. The standard search as defined by the National Library of Medicine is one requiring less than a half-hour of Specialist time, less than 20 minutes connect time, and less than five pages of printout. Users at M.I.T. who desire a more extensive MEDLINE search will be charged at regular rates.

Cost recovery is an essential and fundamental part of the M.I.T. pilot operation. These services must be paid for. While some NASIC host institutions may choose to absorb part or all of the costs and not charge users directly, M.I.T. has chosen to attempt direct recovery of costs. This policy has been adopted in part because the only sure way to test user receptivity of direct charges is to actually charge appropriate fees. We expect that many M.I.T. users will use contract or grant monies to pay for NASIC services and operational use bears this out. More significantly, other users have paid out-of-pocket for services. However, there are potential users, notably undergraduate students, who generally will not have access to research funds. We are currently exploring mechanisms for providing funds to cover the cost of service to these people. One possible model, for example, of sources of funds for student use of NASIC goes along lines similar to dollar support in many places of student use of computational facilities.

A discussion comparing the assumptions underlying the user charges for service with our actual experiences appears later in Task 12, Monitoring and Analysis of Service Operations.

Billing and Accounting Mechanisms (Task 6E)

Another large effort was invested in the development of a mechanism to handle the billing and accounting activities of the initial NASIC pilot service. The existing services and facilities of the M.I.T. Accounting Office are being used as much as possible. The Accounting Office is handling the financial transactions associated with NASIC service and they interface, for billing purposes, between NASIC and the user. Accounts representing each library site plus the NASIC Coordinator's Office have been set up through the M.I.T. Libraries and, for each account, reports by object class for different categories of income are furnished to the NASIC Coordinator by the Accounting Office. However, while M.I.T. retains control over the financial interface between the local NASIC service and the charges to and payments from its users, NEBHE retains control of the financial interface between NASIC services at M.I.T. and outside search services. All bills from outside agencies for services purchased by NASIC at M.I.T. are sent to NEBHE for payment. NEBHE, in turn, bills M.I.T. through the Coordinator's Office for the cost of such services plus a surcharge equal to any discount received by NEBHE from the supplier. NEBHE is apprised of external services purchased by users of NASIC at M.I.T. by either copies of orders and/or reports sent periodically to NEBHE by the NASIC Coodinator's Office. These reports include industrial usage of services; the industrial surcharge collected by M.I.T. is passed along to NEBHE.

This initial phase mechanism could serve as a model for NEBHE in setting up service sites at other institutions, whereupon the local institutional accounting office handles the billing from those NASIC sites to users at those institutions. NEBHE would act as a collection agency only when the local institution could not.

The initial phase operation is currently being tested in practice and the general flow of materials and information is described below. The initial billing and accounting operation functions around a requisition and a word order.

1. At M.I.T., users may pay for NASIC services in several ways:

   a) by authorized requisition against an internal M.I.T. account

b) by purchase order from an external organization

c) by cash receipt showing a deposit made to a NASIC account with the M.I.T. Bursar

d) by personal check

e) for M.I.T. users only, personal billing.

In all cases, the user is given an estimated cost of service, based upon an hour appointment, at the time he arranges his appointment.

2. At the appointment, the user presents his requisition, purchase order, or cash receipt to the Specialist, if payment is by one of those modes. Then, the user and Information Specialist together accomplish a work order based on the user's information problem. At the completion of service, the work order contains the cost to the user for services rendered and the user receives a copy for his own records. The user is also told of the nature and rate of any other costs, such as for off-line output, that are to be billed to him later. The work order indicates whether industrial rates apply. (See Figure B8.)

3. The original work order is forwarded together with any requisition or purchase order to the NASIC Coordinator's Office. The Assistant to the Coordinator prepares from it either a clean order for an off-line external service, or a summary report of any on-line services already rendered. A clean order is sent to the external agency providing the offline service. Copies of a clean order and/or a summary report are sent to NEBHE as notification of services purchased externally by M.I.T. on accounts maintained by NEBHE with those agencies for M.I.T. use.

4. When NEBHE receives an invoice from an external agency, NEBHE makes all payments to the agency and forwards the bill for those services together with any NEBHE surcharge to the M.I.T. NASIC Coordinator's Office. The Coordinator's Office verifies and forwards the bill to the M.I.T. Accounting Office for payment to NEBHE.

5. In the meantime, a copy of the user's work order, with object classes entered against services by the Coordinator's Office, is forwarded to the M.I.T. Accounting Office. All supplementary bills,

such as for off-line output, are prepared by the Coordinator's Office and forwarded to the Accounting Office. The Accounting Office charges the user's account and handles all collections, account transfers, and refunds. The Accounting Office furnishes reports to the M.I.T. Libraries and its NASIC Coordinator's Office.

For MEDLINE services, billing and accounting function as described above except that a) NEBHE is not in the flow and b) bills from the National Library of Medicine are received directly by M.I.T.

As an aid in the computation of charges, Specialists have rate sheets showing the cost by minute for on-line connection to each data base, and for Information Specialist time, and the cost by page or citation for off-line printout. A rate sheet for CA Condensates is shown in Figure B12.

Data Gathering Procedures for Management Information, Qualitative Evaluation, and Statistical Summaries (Task 6F)

The initial operational procedures described in this report have been set up with monitoring and evaluation of services in mind. Data is being gathered in several ways for input to an analysis effort (see Task 12 below). A number of forms have been prepared to capture data during or immediately after an event. Several of the forms are included in Appendix B of this report. Development of the initial data gathering procedures is near completion.

Because of the centralized appointments mode, NASIC publicity names the Coordinator's Office as the place to contact for all inquiries. An inquiry data sheet has been prepared to capture, for example, data about the caller, whether by phone, by mail, or in-person, his location and status, the nature of the inquiry, and the responses given. If an appointment is made, then date, time, and place are noted along with anticipated services, payment method, and problem title. The inquirer is always asked about how he learned of NASIC services. If no appointment is set up, an attempt is made to ascertain the reason. If the inquirer is interested in a data base or service not yet available, his name, location and interests are entered into a special file so that we know what these are and we can personally notify him when such services, or closely-related ones, are made available. In addition, some of these users can aid us in experimentally testing new services under consideration. (See Figures B2, B3, and B4.)

Many inquiries will not come initially to the Coordinator's Office. Many will first be received elsewhere within the library system. The entire library staff is being asked to refer inquiries to the Coordinator's Office or to the Information Specialist in that library. Printed post cards (Figure B5) have been distributed to help accomplish this referral during non-busines hours. Most importantly, an orientation program for the library staff is nearing completion so that the staff (a) may be in a more knowledgeable position about NASIC (particularly the reference staff), (b) can identify with NASIC as a major library activity, and (c) can help NASIC reach a wider audience.

A user who makes an appointment is sent a reminder (See Figure B6) as to time and place, and, if it applies, to bring a requisition or purchase order. The user also receives a problem statement form and is encouraged to complete it and return it before his appointment. (See Figure B7.)

The problem statement form is modelled after those typically used by off-line search centers. The objective is to capture a natural language description of the problem and to define its boundaries. This statement is exceedingly useful if properly completed because the user has then thought about his problem beforehand. If it is returned in advance of the appointment, the Specialist will also have reviewed the problem. The user problem statement serves both the Specialist and the user as a common ground, a point of departure for further probing in working out search terminology and search strategies. The narrative portion of the user problem statement is the most important part. It is essential that the user not attempt to structure his problem at this stage by guess or preconception in anticipation of how material relevant to his problem is indexed. The narrative is more meaningful than lists of phrases or words because it provides additional context by interrelating the phrases. It provides essential details typically absent in search titles. The remaining portions of the user problem statement are useful for the Specialist in at least two ways: (1) it provides additional handles useful in interpreting search feedback, and (2) it can aid in increasing search precision (more limited results) should that be necessary. The user problem statement is also useful in later analysis of the search session and its effectiveness.

The work order captures the essentials of the services rendered to the user. It contains the specific time length, charge data, and method of payment for a user session. The charges to the user are classed by object code for accounting reports. This data is essential to determining the cost recovery effectiveness of the pricing algorithm. (See Figure B8.)

An appointment session log (Figure B9) allows the specialist to capture in free form important decisions and problems that arose during the session. Capture of details of technical problems serve as a basis for securing credits form external agencies as well as providing them with feedback on the quality of the services they are selling.

Off-line outputs are sent to the Coordinator's Office so that (a) user charges can be calculated and billed if these charges were not previously included on the user's work order, (b) identification numbers appearing on the printout can be deleted as a precaution against unauthorized use by others of these numbers, (c) a legend for interpreting coded data on the printout can be enclosed, (d) a thank you for using NASIC can be attached, and (e) an evaluation form covering services and output can be enclosed for subsequent return by the user. As of the date of this report, the legend and the evaluation form have not yet been completed. When full document delivery service is established, a mechanism integrating it with search output will be developed.

The Information Specialists are keeping detailed time breakdowns of their NASIC activities. This information is essential not only for contractual committments, but also for understanding the nature of an Information Specialist's job, and for determining costs of activities. (See Figure B10.)

The Information Specialists also keep a terminal log for all log-on connections and printout requests, whether for training, for experimentation, or for operational services. This data is used in verifying the external agency invoices, and for determining costs of activities beyond that of a user search session proper. (See Figure B11.)

The data that is gathered must be reduced and analyzed. Our initial work on analysis is presented later in Task 12, Monitoring and Analysis of Service Operations.

Marketing and Publicity (Tasks 7 and 10)

Plans for marketing and the development of promotional materials constituted a major effort and it was carried forward in cooperation with the NEBHE staff.

Early on, the characteristics of the major channels of publicity at M.I.T. were identified. In addition, statistics were gathered about the M.I.T. population by department and laboratory and by status, as, for example, faculty, students, staff, visitors. Information already compiled at the Institute in the form of directories to the major research interests of the M.I.T. community was also collected. The population and research interest data has been used in the selection of data bases and in the determination of selective mailing lists by department. The publicity channels primarily used to-date center around news releases, brochures, mailings to selective identifiable groups, meetings, and demonstrations.

Several brochures specific to NASIC at M.I.T. were prepared and added to an earlier brochure produced by NEBHE describing the general objectives of NASIC. One of the new brochures for M.I.T. gives an overall picture of the initial NASIC services available at the Institute. This general brochure is complemented by three others, each containing important general information but each also contains content specific to a different data base--one in chemistry, one in education, and one in business. A fourth brochure, prepared and produced at M.I.T. expense, covers medicine. A price list complements all the brochures. The price list is illustrated in Table 2. The brochures are illustrated in Appendix C.

The publicity events included, in chronological order:

1. An initial news release to Tech Talk, the M.I.T. Community weekly newspaper, appearing 12 September 1973 and describing the NASIC subcontract effort at M.I.T. This release coincided with the beginning of the Fall semester. The release was issued to the Associated Press and was picked up by at least the following national and local newspapers: The New York Times (23 September 1973), the Worcester, Mass. Gazette (20 September 1973), and the Fall River, Mass. Gazette (20 September 1973).

2. At the invitation of Miss Nicholson, Professor Reintjes spoke to the M.I.T. Library Council on 17 October 1973.
3. Miss Nicholson spoke about NASIC before the M.I.T. Faculty Council meeting on 7 November 1973. This was an agenda item for their monthly meeting.
4. A letter prepared by Miss Nicholson announcing the opening of NASIC services on 15 November was sent together with a copy of the General Brochure to all 2600 M.I.T. faculty and staff members on 12 November 1973.
5. The campus student newspaper The Tech interviewed Miss Nicholson and Professor Reintjes about NASIC services and ran their article on 27 November 1973.
6. The letter by Miss Nicholson to faculty and staff and the General Brochure was sent to all 1388 M.I.T. research assistants and teaching assistants on 29 November 1973. One or more brochures on specific data bases were included in the mailings to all 374 graduate students in selected departments. The chemistry brochure was sent to those students in the Chemistry, Chemical Engineering, Biology, Metallurgy and Material Science, and Nutrition and Food Science Departments. The medicine brochure was also sent to students in some of those departments. The education brochure was sent to students in Foreign Literature and Linguistics. The business brochure was sent to students in Economics.
7. An article about NASIC services appeared in Tech Talk on 5 December 1973.
8. William Duggan and Alan Benenfeld spoke before the M.I.T. Administrative Officers regular monthly meeting on 20 December 1973. The billing and accounting interface with NASIC services was emphasized.

9. A letter sent by Miss Nicholson to each library staff member about the opening of NASIC services was a prelude to a series of orientation meetings held beginning in December by the NASIC Coordinator and the Information Specialists with the staff of each library. The meetings were to familiarize the staff with NASIC services and to further encourage development of effective communications and coordination between traditional library services and computer-based library services. Considerable emphasis is placed on NASIC as an integral part of library services.
10. A Chemical and Engineering News reporter was referred to the NASIC Coordinator by the Tech Talk Office. A brief announcement appeared in the 24 December 1973 issue of Chemical and Engineering News. Some inquiries by non-M.I.T. people were received as a result of that announcement.
11. A series of six online demonstrations and one seminar plus demonstration about NASIC services were held in January during M.I.T.'s Independent Activities Period (IAP) by Mary Pensyl and the Information Specialists. There were two demonstrations of the education data base, one of chemistry, two of business including one at the seminar, and two of the medical data base. The demonstrations consisted of 10 minute introductory talks, a 20 minute prepared search to illustrate features of the data base, and a 10 minute question and answer period. The seminar consisted of a 30 minute discussion followed by a 20 minute prepared search and a 10 minute question and answer period. Approximately 120 people attended the demonstrations and about 20 people attended the seminar. These activities drew a mixture of M.I.T. faculty, students, and staff, in addition to a number of non-M.I.T. personnel, particularly librarians from other schools in the area. We know of three appointments that resulted from this effort. One M.I.T. department began to investigate the possibility of setting up funds for its students to draw upon for NASIC services.

12. A two-day Symposium was held on 7-8 February 1974 at M.I.T. and jointly sponsored by the M.I.T. Electronic Systems Laboratory, the M.I.T. Libraries, and the New England Board of Higher Education. The symposium covered two topics: (1) reports on the results of Project Intrex, and (2) reports on the status of the Northeast Academic Science Information Center Invitations were extended to the Deans of the fourteen graduate library schools in the northeast and to the Heads of 43 university, college, and other large research libraries, also all in the northeast. Each institution was invited to send two representatives. Sixty persons attended. Natalie N. Nicholson launched the NASIC portion of the program with an "Introduction to NASIC at M.I.T." on Thursday evening. Professor Reintjes introduced the Friday program which included presentations on "The Background and Objectives of NASIC" by David Wax, "A Rationale for NASIC at M.I.T." by Richard Marcus, "The Development of NASIC at M.I.T." by Alan Benenfeld, "The Integration of NASIC into the M.I.T. Libraries" by Mary Pensyl, and the "Future Plans of NASIC" by David Wax. The program concluded with demonstrations by the Information Specialists of the NASIC facilities in the M.I.T. Libraries.
13. On 11 February 1974, the Lincoln Laboratory Library Director, Mr. Lloyd Rathbun, was sent several hundred copies of each of the NASIC brochures and price list for distribution to Lincoln personnel. An announcement about NASIC appeared in the 15 February issue of the Lincoln Laboratory Library Scanner. A Lincoln Laboratory bibliographer, Ms. Sara McNeil, was named as a contact person for Lincoln staff. Lincoln has since issued a purchase order through their Library for NASIC services.
14. The letter by Miss Nicholson was revised and it and the General Brochure were sent on 15 February 1974 to all 767 staff at the Charles Stark Draper Laboratory, Inc.
15. The M.I.T. Industrial Liason Office in mid-February contacted by telephone about 20 firms in the Greater Boston area about the availability of NASIC services. This was a follow-up to a previous letter by ILO announcing that services to the industrial community would be forthcoming.

16. On 20 February 1974, David Wax and Alan Benenfeld spoke before the Special Libraries Association, Boston Chapter, Science-Technology Committee. The development of NASIC and its implications to special libraries were discussed. About 80 librarians from the Greater Boston area attended. Several inquiries from attendees at that meeting have since been received about providing services to particular libraries. We are aware of at least one announcement, in the Massachusetts College of Pharmacy Library's bulletin, of NASIC services.
17. On 5 March 1974, a demonstration and discussion about NASIC was held at M.I.T. with 8 librarians from the University of Massachusetts, Boston. Information about NASIC services has since appeared in their library bulletin.
18. On 10 March 1974, a revised letter by Miss Nicholson was sent along with copies of brochures and price list to Miss Helen Brown, Director of the Wellesley College Library for distribution to Wellesley faculty and staff. M.I.T. and Wellesley have a cross-registration program and other reciprocal agreements.
19. On 14 March 1974, Mary Pensyl gave a presentation on NASIC services in the M.I.T. Libraries at a joint meeting of the Harvard and M.I.T. Library Staff Associations. About 60 persons attended.
20. On 1 April 1974, Mary Pensyl spoke on NASIC as an M.I.T. Libraries service to 30 persons attending a meeting of the M.I.T. Women's Forum.

The service site in the Science Library was strikingly repainted by the M.I.T. Libraries. Consideration is now being given toward a similar effort at some of the other service sites to enhance visibility and command attention. Plans for effective displays of brochures at each site are underway as are plans for poster displays.

Word-of-mouth advertising by satisfied customers about good, efficient, and effective NASIC service is a most desireable goal. To this end, the Information Specialist training program and the library staff orientation meetings have been particularly sensitive to furthering the initial enthusiasm, espirit-de-corps and cooperation among all participants, backed up by the development of coordinated procedures for effective service.

Our experiences to date indicate that all publicity mechanisms will yield some response but that word-of-mouth advertising is gaining in predominance. As Phase 2 gets underway, consideration is being given to new avenues of publicizing services, in particular, personal contact with potential prime user groups. We also plan in Phase 2 to survey both users and non-users of NASIC services. An initial evaluation of the marketing effort and some plans for further effort are given later in Task 12, Monitoring and Analysis of Service Operations.

Testing of Service Procedures (Task 9)

During the two week period preceding the opening of services, dry runs on parts of the operation were held. These were concentrated entirely on interviewing and performing searches with trial users. Testing of all other procedures is being done under real conditions of service as part of the continuing monitoring and analysis of the operation (see Task 12).

Initiate Service (Task 11)

This task represents a major milestone in the history of NASIC. On Thursday, 15 November 1973, NASIC at M.I.T. officially began services by taking appointments for service beginning Monday 19 November. Interestingly enough, inquiries were received at the Coordinator's Office on 14 November from faculty and staff who had already begun to receive the announcements sent to them. Service operations were launched just four months after the start of M.I.T.'s subcontract. Task 12 below summarizes and analyzes the activities following the start of a service operation.

Monitoring and Analysis of Service Operations (Task 12)

The discussion of each of the preceding tasks has been mainly a descriptive report of the design, development, and implementation activities associated with NASIC services. In this task we look at the service operations themselves and compare design with practice. Feedback from monitoring and analysis of operations is vital not only to making changes to the operation or to the policies governing the operation, but also to understanding better the functions being performed and the needs of the user community. Because we have been on-the-air for only a limited time, feedback has been useful so far mainly to improve understanding and to modify general plans; few changes of consequence have been made as yet to either policy or to operation. The following analysis draws upon the statistical data appearing in Tables 3 through 9, upon the descriptive reports above, and upon user receptivity reports.

It may be helpful to begin with a statistical characterization of use. All statistics refer to the period 15 November 1973, the date services began, through 28 February 1974.

There were 57 users; of these, 29 used one of the three NASIC data bases and 28 used the MEDLINE data base. All searches were online retrospective searches. There were no current awareness searches, either online or offline, and no offline retrospective searches. All searches but one were run on an appointment basis with the user present. The one exception was a second search for a user run in a delegated mode using the INFORM data base. Table 3 shows the breakdown of users by data base and by the library in which service was received. Table 4 shows the breakdown of users by organization and status for each data base. For M.I.T. campus users, Table 5 shows for each data base the distribution of data base users by department or laboratory. Table 6 shows the publicity mechanisms to which M.I.T.-affiliated users responded. Table 7 shows the methods of payment selected by all searchers distributed according to their status. The data in Tables 6 and 7 do not distinguish between NASIC data bases and MEDLINE because the data base was not expected to influence the distribution.

Table 3

NASIC AT MIT

SUMMARY DATA

15 NOVEMBER 1973 TO 28 FEBRUARY 1974

NUMBER OF SEARCHES: 29 NASIC, 28 MEDLINE

(ALL ARE ON-LINE RETROSPECTIVE SEARCHES)

SEARCH LOCATION AND DATA BASE:

	CHEM	ERIC	INFORM	NASIC TOTAL	MEDLINE
BARKER	9	1	--	10	--
DEWEY	--	2	4	6	--
HUMANITIES	--	1	--	1	--
ROTCH	--	--	6	6	--
SCIENCE	5	1	--	6	28
OTHER LIBRARY	--	--	--	--	--
OFFICE/LAB	--	--	--	--	--
TOTAL	14	5	10	29	28

Table 4

NASIC AT MIT

SUMMARY DATA

15 NOVEMBER 1973 TO 28 FEBRUARY 1974

USER AFFILIATIONS:

	CHEM	ERIC	INFORM	NASIC TOTAL	MEDLINE
MIT/CAMPUS TOTAL	12	4	8	24	25
FACULTY	5	--	--	5	6
GRADUATE STUDENT	2	2	5	9	11
UNDERGRADUATE	--	--	--	--	4
OTHER STAFF	5	2	3	10	4
MIT/LINCOLN	--	--	--	--	--
DRAPER	--	--	--	--	--
WELLESLEY	--	--	--	--	--
OTHER UNIVERSITIES (TOTAL)	2	1	2	5	2
FACULTY	--	--	--	--	--
GRADUATE STUDENT	2	1	2	5	1
UNDERGRADUATE	--	--	--	--	--
OTHER STAFF	--	--	--	--	1
GOVT. AGENCIES	--	--	--	--	--
INDUSTRIAL/COMMERCIAL	--	--	--	--	1
OTHER AFFILIATIONS	--	--	--	--	--

Table 5

NASIC AT MIT

SUMMARY DATA

15 NOVEMBER 1973 TO 28 FEBRUARY 1974

DEPARTMENT AFFILIATIONS OF

MIT/CAMPUS USERS:

DEPARTMENT	CHEM	ERIC	INFORM	NASIC TOTAL	MEDLINE
BIOLOGY	1	--	--	1	1
CHEMICAL ENG.	1	--	--	1	--
CHEMISTRY	2	--	--	2	1
ELECTRICAL ENG.	--	1	1	2	1
MATHEMATICS	--	--	--	--	1
MECHANICAL ENG.	2	--	--	2	5
METALLURGY	1	--	--	1	1
NUCLEAR ENG.	1	--	--	1	--
NUTRITION	--	--	--	--	5
OCEAN ENG.	--	--	1	1	--
SLOAN SCHOOL	--	1	4	5	1
URBAN STUDIES	--	--	--	--	2
HEALTH SCIENCES PROG.	2	--	--	2	4
LIBRARIES	--	1	--	1	--
MAGNET LAB.	1	--	--	1	1
OASIS	--	1	2	3	--
PLANNING OFF.	--	--	--	--	1
SEA GRANT	1	--	--	1	1

Table 6

NASIC AT MIT

SUMMARY DATA

15 NOVEMBER 1973 TO 28 FEBRUARY 1974

PUBLICITY RESPONSE OF MIT-AFFILIATED USERS: (Combined NASIC and MEDLINE)

	FACULTY	CAMPUS GRAD.	UNDERGRAD.	OTHER	LINCOLN LAB	DRAPER LAB	TOTAL MIT
MAILINGS (LETTERS AND/OR BROCHURES)	3	4	2	3	--	--	12
THE TECH ARTICLES	--	2	1	1	--	--	4
TECH TALK ARTICLES	--	2	1	--	--	--	3
COLLEAGUE	1	6	2	1	--	--	10
DEMONSTRATION	--	4	--	1	--	--	5
LIBRARY STAFF REFERRAL	2	2	--	--	--	--	4
DISPLAY/OBSERVER IN LIBRARY	1	5	1	2	--	--	9
*POSTER/DISPLAY OUTSIDE LIBRARY	--	--	--	--	--	--	0
REPEAT USER	3	--	--	4	--	--	7
OTHER SOURCES	--	--	--	1	--	--	1

*Not implemented in the reporting period.

Table 7

NASIC AT MIT

SUMMARY DATA

15 NOVEMBER 1973 TO 28 FEBRUARY 1974

METHOD OF PAYMENT FOR SERVICES: (Combined NASIC and MEDLINE)

	MIT-USERS						NON-MIT USERS	
	FACULTY	GRAD.	UNDER-GRAD.	OTHER	LINCOLN LAB	DRAPER LAB	ACAD.	COMM.
MIT REQUISITION	11	15	1	12	--	--	1	--
PERSONAL CHECK	--	4	3	2	--	--	6	--
CASH	--	1	--	--	--	--	--	--
PERSONAL BILL THRU MIT	--	--	--	--	--	--	--	--
PURCHASE ORDER	--	--	--	--	--	--	--	1

Table 8

NASIC AT MIT

SUMMARY DATA

15 NOVEMBER 1973 TO 28 FEBRUARY 1974

CHARACTERISTICS OF ON-LINE SEARCHES BY APPOINTMENT:

	CHEM	ERIC	INFORM	NASIC OVERALL	MEDLINE
SEARCHES BY APPOINTMENT:	14	5	9	28	28
AVERAGE ADVANCE TIME IN ARRANGING AN APPOINTMENT: (business days)	6.4	3.2	5.8	5.7	5.2
AVERAGE CONNECT TIME: (minutes)	42	36	29	37	47
AVERAGE APPOINTMENT LENGTH (minutes)	77	62	76	70	71
AVERAGE RATIO CONNECT TIME TO APPOINTMENT TIME:	.55	.58	.38	.53	.66
APPOINTMENTS WITH A MACHINE PROBLEM:	3	0	0	3	3
AVERAGE TOTAL PROBLEM TIME: (minutes)	14	--	--	14	13
OFF-LINE PRINT REQUESTS:	7	4	5	16	20
AVERAGE OUTPUT (pages):	44	44	28	39	50
(citations):	211	110	87	131	394
AVERAGE COMPUTER PLUS ADMINISTRATIVE CHARGE: (before any allowances)	$39.52	$26.95	$32.22	$34.90	$13.86
AVERAGE SPECIALIST CHARGE: (before Introductory Credit)	$10.25	$ 8.22	$ 8.58	$ 9.36	$ 9.32
AVERAGE PRINTOUT CHARGE:	$ 6.12	$ 5.33	$ 7.18	$ 6.21	$ 4.53
AVERAGE USER COST: (assuming offline printouts requested and no credits given)	$55.89	$40.50	$47.98	$50.47	$27.71

It is helpful to characterize further the search sessions within a search mode. To date, we have had essentially only one mode, an on-line search by appointment. Table 8 characterizes quantitively the features of the 28 NASIC data base searches and the 28 MEDLINE searches that were held by appointment and with the user present. The charges vary widely among the data-bases. The amount of offline printout also appears to be influenced by the size and the number of years covered by the data base. On the other hand, other parameters, which we shall examine, show little difference in usage among the data bases. Some of the fluctuations we do see may be attributed at this early stage to the one or two users at the extreme for each data base who have influenced the average values reported here. For example, the 394 citations printed offline per average session for MEDLINE is heavily weighted by two users out of 28 who received 1431 and 1338 citations respectively, excluding those users, the average MEDLINE printout contains 282 citations. This is still a large number but the medical data base is quite extensive and the figure is comparable to the more than 200 average citations printed for users of the similarly extensive chemistry data base. With a larger number of users, extreme cases will have less influence on the average values--or, indeed, we may discover that these "extremes" are not really abnormal.

The appointment itself runs some 70 minutes, of which 37 minutes, a little more than half the time, is spent in an actual on-line connection to the retrieval system, in our case either the SDC ORBIT or the NLM or SUNY MEDLINE systems. About 10% of the searches have encountered more than minor machine connection problems, occasional disconnect or line noise problems are more frequent but they are considered minor if recovery is almost instantaneous and if the search is unimpeded. If machine problems impede the search in any way, the user is given credit against his total charge for the amount of time involved.

Only 50% of NASIC users and 60% of MEDLINE users request offline printouts. The request is issued online during the search session but the printing is done offline on a high speed printer and sent by airmail. For users who have a lengthy list of references, it is often cheaper to obtain offline printouts rather than by printing online at 30 characters per second.

However, users do obtain enough particulars online so that their work can proceed until the printouts are received a few days later (on rare occasions, the next day). The average offline request runs some 40 to 50 pages. The number of citations varies with the comprehensiveness of the data base. The reason for at least some users not issuing an offline print request is that no relevant material has turned up in the search process. These negative searches are not at all uncommon or unexpected among research-oriented users who are often simply seeking reassurance that no one else has done what they propose. Although we have not specifically analyzed for the number, the percentage of negative searches is probably at least 20 or 25 per cent and may run as high as 40 or 50 per cent.

The average total charge to a user is $50.47 for searching a NASIC data base and $27.71 for searching MEDLINE. This is the charge assuming offline printouts are requested and without making any credit allowances. In this period, M.I.T. users receive an introductory credit for the Specialists's time and so the real cost to an M.I.T. user has been about $9 less in each case. The average cost of the printout component is $6.21 for NASIC data bases and $4.53 for MEDLINE. The MEDLINE printout is a per-page charge. The NASIC printout cost has been confounded by the change from a per page to a per citation charge as discussed under Task 6D, Service Charges. Since the per citation rate was in effect only for February, we expect the average NASIC printout charge to rise. The major component of the total cost to a user is the charge covering computer connection, computer searching, and M.I.T. administrative costs (cf. Task 6D, Service Charges). For NASIC the average charge for this component is $34.90 and for MEDLINE it is $13.86. The MEDLINE cost is considerably lower because the National Library of Medicine subsidizes entirely the computer search cost and because the connection cost is lower than the connection costs to SDC.

What affect does price have on data base usage? This is perhaps the most important question we can address. The availability of the subsidized MEDLINE system to the user community through the same channels as the NASIC data bases gives us some additional handles with which to begin answering the question. At this point in time, there is no definitive answer because price is one of many confounding factors that may motivate a person to use computer-based reference services such as NASIC and MEDLINE. For example, price, marketing, need, prior familiarity with data base, availability of funds, complexity of the search, urgency of results, convenience, and influential or peer users, are all candidates to influence at least some users. Price is the easiest to measure and study but we do not yet know whether it is the most important of all factors or only one of several important factors. There may be threshold effects associated with price or other factors which influence usage. What can we tell at this time from our initial data?

There is no initial onslaught of users for any data base. The number of MEDLINE users runs neck and neck with the total number of NASIC data base users. However MEDLINE is used twice as heavily as the most frequently used NASIC data base in chemistry. Price could explain the difference between CHEMCON and MEDLINE usage but other factors may be at least as, or even more, influential. One such factor is the proportion of the total community interested in a data base. For example, the chemistry data base, CHEMCON, is used three times as often as ERIC, and the INFORM business base is used twice as often as ERIC. Both CHEMCON and INFORM are more expensive than ERIC, but more people at M.I.T. are active in chemistry research than in management research, and more in the latter than in educational research. Thus for these three areas, usage ranks in order of the size of the interested population, and not inversely with cost. What about interest in the medical research area? At M.I.T., interest in medical and health care research is known to be highly diffused throughout almost all departments. This is confirmed by the distribution of users shown in Table 5. Interest in medical science and technology and related health care systems is at an all time high nationally and at M.I.T. These are areas where research funds are more fluid but there is also a genuine trend at M.I.T., as elsewhere, to turn the attention of more of its research resources to social and environmental problems. Thus the greater usage of MEDLINE may also result from and be proportional to the number of people having interest in it rather than it being an effect of price, even when the price is heavily subsidized. The explanation based on interest is consistent not only between NASIC and MEDLINE but also between the respective NASIC data bases. One way to test this hypothesis is to compare the ratios of data base use to the population ratios of fields of interest of researchers. We hope to tackle this in phase 2 of our work but the test is complicated by the fact that today's highly interdisciplinary research shows little respect for departments, laboratories, and even data bases initially organized and named along lines of more traditional disciplines.

If interest can explain data base use at least as well as price, are there results for which price is more clearly a major agent of influence? Analysis of Tables 4 and 7 sheds considerable light on this area. The status of our users appears to be influenced by the price. Of 49 M.I.T. campus users (24 for NASIC data bases, and 25 for MEDLINE) only 4 have been undergraduates and all 4 used MEDLINE. Undergraduates and graduate students each represent about a third of the M.I.T. professional body, the faculty about ten percent, and the research and administrative staffs the remainder. Thus undergraduate use of MEDLINE is still less than proportional to their population. On the other hand graduate students have used all data bases, and except for chemistry, in greater proportion than their population. More graduate students are involved in research than undergraduates. But even more significantly, graduate students have greater access to contract and grant monies than do undergraduates. As Table 7 shows, 3 of 4 undergraduates paid out-of-pocket (by check) for their searches, but only 25 percent (5 of 20) of the graduates paid out-of-pocket. Table 4 shows faculty and other professional staff (research and administrative) use. Other staff use has been much higher than expected. From Table 7, all faculty users charged services against contract funds, as did 6 out of 7 other staff people. Thus, price appears to influence the class of user.

If this is true, then it becomes necessary to find funding mechanisms to pay, partially or completely, for searches by an undergraduate or any other potential user who lacks personal funds and has no recourse to grant monies. This area is currently being addressed at M.I.T. The problem is similar to the one of providing funding for undergraduate use of computational facilities.

Is there any other discernible major influence of price? Price appears to influence the type of service users seek. To date we have had no use of offline services. A review of Table 2 readily shows that the offline services currently offered are more expensive to use than online services. The offline prices are for search periods of one-year of a data base. In the online mode, the cost is per connect-hour and it is possible to search several years of data base coverage at one time. The connect times in Table 8 show that the average online search is accomplished in well under one hour of terminal connect time. Terminal connect time would have to increase by an extraordinary percentage to exceed the cost of a comparable offline search. While we have not yet performed online current awareness searches, there is no question that the above argument on relative cost still remains valid. Our users have recognized these differences in relative cost and some of them have made explicit comments to that effect. But the issue is also confounded by questions of convenience. There is no question that online searches give the user faster results. The searching process is highly interactive. There is immediate feedback. The user obtains results as his search proceeds. There is no delay except for additional extensive printouts (the delay in mailing is only one-way because the search itself has already been performed). Thus, online services are not only lower in cost, they are also more convenient. It is worthwhile re-emphasizing at this point that NASIC represents services available on systems maintained by others whose costs are shared widely. The commonly held view in many quarters that online systems are more expensive than offline systems just does not carry over to the kinds of computer based reference services that NASIC activities represent.

Does price affect the way in which a data base is used? It appears from the data in Table 8 and the above discussion of that table, that the characteristics of searching a data base, at least in the appointment mode, are not influenced by price. Users do not search a data base for a longer time simply because it costs less money. If they did, the average connect time for each data base should show a progressive increase with decreasing cost but it does not (cf. Table 2). It is much more likely that the size and comprehensiveness of the data base being searched influence the connection time. Indeed, the bigger the data base, the greater the connection time. INFORM is as expensive as CHEMCON, but it is the smallest and least comprehensive data base; it has the lowest connect time. The other side of the coin, namely the time the user spends with the Specialist in his appointment but not in an online connection, is not as amenable to interpretation. We suspect that this time, which is largely reviewing the problem and setting up an initial strategy, may be a function of the kind of indexing, the vocabulary controls, and the retrieval aids that are associated with each data base. Much more study is required in this area, particularly because it may turn out that other factors limit the total time available for both pre-search strategy plus on-line search, and that these limits, in turn, influence the proportional breakdown. Some factors that might influence the total time limit on pre-search plus search are psychological conditioning of Specialist and user to expect about one-hour, anxieties about exceeding an appointment block, and external boundary conditions such as having another commitment scheduled, perhaps because of prior expectations of appointment length.

Why hasn't there been greater use of computer-based services during this 2 1/2 month period? One might suspect that one reason was timing. Thanksgiving, fall semester finals, Christmas recess, the three-week Independent Activities Period in January, and a recess between IAP and the spring semester all fell within this period. However, business did not suddenly boom once the second semester began in early February. Three factors much more important than timing may provide the answer: (1) paying for services, (2) data bases of interest, and (3) marketing. Earlier, we said that price alone cannot explain which data bases are more frequently used. However, the fact that users are asked to pay for the services they receive, services that traditionally have been "free," may be a psychological, as well as an economic, barrier. Potential users who come to NASIC have already overcome the barrier. Additional marketing will help others over the barrier. Once that barrier is overcome, the particular data bases of interest to a user will be searched, at least as long as their absolute cost is not unreasonable in terms of value received. The people who come to us want service. The relative cost of different bases may be less important to potential users than the acceptance of having to pay for service in the first place. At present, direct payment for information (copying services excepted) is new to most people. But administrators of information sources know all too well that their cost is a cost actually borne by users indirectly through institutional overhead or by taxes. We will not get into the question of whether the cost of all information services should be borne directly or whether some or most such services should continue to be paid for indirectly. But shortly we will address that question with respect to computer-based literature searches. First, let's consider whether there is a value to having these services. If there is none, cost would be moot.

Almost all of these computer-based services have a printed counterpart such as Chemical Abstracts for CHEMCON or CA-Condensates, Index Medicus for MEDLINE, Research in Education for ERIC. In fact, the machine readable data base is really a by-product of producing the printed form. What then are the values gained by working with the machine version that are present to a lesser degree or absent when working with the printed publication? The major values of a machine-based search are:

1. Less time required by a user to do a search, especially if the search is by subject.
2. Only one search need be performed for the cumulated years of coverage of the data base.
3. Physical manipulation of multiple volumes is avoided.
4. Complex logical combinations can be used that are difficult or impossible to carry out in manual mode.
5. Typically, the indexing and retrieval mechanisms go beyond those present in printed form so that there is greater accessibility to each item (the number of access points is increased).

6. Computer printout generally eliminates the need to take notes or copy citations.

Time saved and accessibility are the two most important. Unfortunately today in academic environments, people's time is less highly valued than say, in industry, but this is changing. There is greater emphasis on productivity in the office, laboratory and classroom and this will, in turn, influence research in the library. Accessibility to records is improved by deeper indexing and by machine techniques but it requires particular knowledge, understanding, and training to be used properly. Users of machine services, by so doing, can improve their skills in conducting searches in any medium and thus their use of the library in general can become more effective. This is very important because machine services are in no way a replacement of traditional services; each type of service complements the other.

We have no doubts about the value of machine-based services to users. Do users have doubts? Because of the small number of users so far, it might seem that potential users do. However, our initial actual users have been highly enthusiastic about the services they received; more on this later. What then is the problem? For one thing, it takes time for a new service to make itself known and to build up a consumer group. We began our operations with only four data bases. We need to make available a larger number of data bases of interest to more segments of the community. We are proceeding early in Phase 2 to train Specialists in working with additional data bases. In two years we expect to have about twelve to fifteen such data bases available.

But just making more data bases available will not in itself solve the problem. An improved marketing effort is essential in order to develop an initial thrust of interest in each new data base. Over the somewhat longer-range, word-of-mouth should become more and more effective as a source of new users. An initial core of satisfied users need to be obtained (while taking pains to see that initial service is of high quality and effective because it is much more difficult to overcome negative reactions.) The marketing effort to-date has relied on a number of devices, mainly that of written materials (letters and brochures) sent to potential users, demonstrations, and news stories (refer to the lists under Task 7 on Marketing and Publicity). Table 6 shows that all publicity mechanisms yield some response in the M.I.T. community; in addition, a monthly breakdown of this data would show that word-of-mouth by satisfied users is gaining in predominance with time. Within the library, referral by library staff, displays, and demonstrations have been important avenues. Of 49 M.I.T. users, seven have been repeat customers (a form of self-referral and an ultimate test). But the publicity for reaching users who haven't come into the library generally has been passive. We personally know of several faculty and other staff members who received brochures by mail but who, when asked informally about the mailing, have little or no recollection of it. The mailings have gone unnoticed or unread by many recipients. This form of advertising, while having a definite role, is not sufficiently differentiable from all the other mail received by our potential users. A boost is needed to create awareness of NASIC among prime user groups. Personal contact and follow up is a necessity. However, this does not mean a hard sell which probably would not go over well in academia. It does mean a personal touch which can tune into specific needs of the potential user. New publicity mechanisms must relate more heavily to specific and current needs of potential users, such as in thesis work, contract preparation, or course development, to motivate them to become real users. In addition, it is important to emphasize more the values of a machine based search since these values are not necessarily obvious to users. In Phase 2 we plan to undertake a more active publicity campaign. This will include, among other techniques, meeting with faculty individually or in groups, phone-call follow-ups to mailings, and enlisting the aid of satisfied users. Libraries usually have not actively promoted their services. But for a new product such as computer-based searching to get off-the-ground and find acceptance within the community, active promotion is a necessity. It may even attract new users to other library services.

We know from the short-time that we have been on-the-air that there is a demand for the kind of services being offered. That there is some strength to this demand can be adduced from the fact that many users are paying out-of-pocket for such services, although the majority of users to date have access to contract or grant monies. We also know that a greater number of data bases and a more personal marketing effort can be expected to improve the growth of service use. Given the value and the use, who should pay for services such as these? This is a question that each institution must, of course, answer for itself in terms of its own size, budget, and funding mechanisms. To take on an extremely simple and conservative calculation, suppose the expenses associated with a typical search including all direct and indirect costs for computer time, communications, direct-labor, administration, terminals, advertising, and materials, is $75. If a modest 500 searches can be expected in the course of a year, then $37500 is needed to cover expenses; for 1000 searches, $75000 is required. The reader can make multiplications for other demand rates but it should be obvious that a dollar level is required for complete subsidization of even moderate use that could be difficult for an organization to raise in today's economy on top of already seriously strained budgets especially where the cost-effective benefits may not be apparent for this new kind of service. True cost effective calculations will require looking at the user, and the time values of the user, as part of the service system. If a library cannot completely subsidize these services, then initially at least, it may be better to offer services at fees that recover most costs and by so doing, demonstrate to higher administrative levels that, even with a charge, there is an actual demand for these services. Simultaneously, other mechanisms can be sought to support all additional costs through rearrangement of budget priorities, through cost-sharing with departments, or through entirely new avenues. For example, M.I.T. is, at least currently as an introductory offer, subsidizing the time of the Information Specialist with all other costs charged to users. Many other ways to share the costs are possible, and there also are many ways to set up a pricing structure but it is not our purpose in this report to review them. It _is_ our purpose to point out that a new service of value to users and in demand by them, is available and that it can be offered by a library, even if the user must be charged some, if not all, of the cost.

If a library chooses not to offer such services, other campus organizations such as individual departments or a computation facility probably will, and for a fee too. If a library relinquishes this kind of service, it would be, in our view, unfortunate and indeed a disservice to the campus and to the profession for the following reasons:

1. librarians represent the bibliographic expertise available to the community
2. these services are, in effect, a powerful catalog and index to materials available through the library
3. information needs of users require availability and integration of diverse resources, one complementing another
4. separation or isolation of these resources undermines the ability of any one resource to be used effectively
5. the end user suffers because different organizations can each satisfy only a part of his information needs.

Let's turn to other analyses that can be made at this time of the M.I.T. operation. Table 8 tells us that the average number of business days between the time a user arranges for and holds an appointment is 5.7. The median wait is only 3 business days; that is, half of the appointments are held within 3 days of the user's call. The average wait has been affected by several users who found it necessary to reschedule their initial appointments or who otherwise had full calendars and found it difficult to initially arrange an appointment for a time slot when the data base of interest was also up. The waiting time may seem excessive but no user has requested immediate service. It is possible, however, that we may not be hearing from users who want instant or demand service such as they get themselves by searching on their own manually in the library. Nevertheless, we suspect that there is a greater tolerance on the part of most users for waiting for service than is commonly thought. On the other hand, if the median wait approaches a week or more, this is likely to be excessive and beyond tolerance. The tolerance of users may be because they know they will obtain system feedback at the time of their appointment. In our initial operations, a user could receive service for any data base, except MEDLINE, in any divisional library. However, the time of the appointment rather than its location may be more important to a majority of users, although several users have indeed requested service in a particular library. The issue of service facilities geographically convenient to the user is clouded by the SDC three-hour time windows during which only certain data bases are accessible. We hope to clarify the picture of user convenience during Phase 2. We also wish to experiment with providing service directly in the office or laboratory of a user. In addition to time and location of an appointment, some repeat users have specifically requested service with the same Information Specialist who served them earlier; this makes good sense particularly if the same user problem is to be searched against another data base. It also represents the beginning of a professional relationship between Specialist and user.

There is additional data to draw upon which will help to complete the picture of NASIC operations at M.I.T., particularly in terms of background functions and duties. For example, a user who either phones in to, or stops by, the Coordinator's Office spends from 5 to 20 minutes conversing with the Coordinator's Assistant about the services available, their cost, which one(s) are applicable to his problem, arranging an appointment, and the nature of the user problem statement. Two or three subsequent calls are often necessary, but these are generally much briefer, and the nature of these conversations mainly is to confirm an appointment, remind a user of an appointment, or request return of the problem statement.

The interaction that transpires is an essential, even influential, element in the marketing process. A potential user already has been motivated to make an inquiry. Some of these people are prepared to arrange an appointment, others need further reassurance that NASIC services will be helpful to them, and others cannot be served at the present time by the available data bases but may be candidate users of an expanded service. A file of prospective users and their interests is being kept.

The centralized mode for dispersing information and for appointment arrangements has been most beneficial in lessening the burden that otherwise might be carried by the Specialists and other personnel in each library. The Specialists can concentrate more on the actual service than on arrangements for service. In those cases when the first point of contact of a prospective user with NASIC is within the library through a staff member or by seeing a display, the prospect is referred to the Coordinator's Office for more detailed information.

The Specialists spend about ten to fifteen minutes before an appointment reviewing the user problem statement to familiarize themselves with the problem. This may include consulting printed forms of a data base or associated thesauri and other retrieval aids. In several cases, the Specialists have also conducted online pre-appointment searches of about 10 to 15 minutes to test out vocabulary and strategies. These pre-appointment searches tend to be undertaken when the Specialist is less familiar with a specific subject area or its treatment in a data base and this may be considered part of the learning process. We expect that this type of activity will decrease as the Specialists continue to gain experience in providing these services. A natural question to ask is what effect a pre-appointment search has had on the effectiveness of the actual search during the appointment, but we have not yet looked into this, either quantitatively or qualitatively.

After an appointment, a Specialist may spend up to an additional half hour documenting the major events transpiring during the appointment. These "minutes of the session," so to speak, are an important tool for any later analysis of the sessions, the problems (technical or logical), and the reactions. If a user does not wish to keep the print record of his search (a rare occurrence), then these are made part of the documentation. In the absence of that record, the Specialist generally notes the strategy that has been used. These can be most helpful later for training new Specialists, for demonstrations, and for referral from a similar problem by another user.

The total pre-appointment plus post-appointment activity of a Specialist may be, on average, about an hour or almost as long as the average appointment itself. Some of the pre-appointment time is a result of inexperience. A good portion of the post-appointment is a direct result of the experimental mode in which M.I.T. is studying and testing these kinds of services for NASIC.

None of the costs associated with the Specialist's pre-appointment or post-appointment activities are charged directly to users. The pricing structure, which was described in Task 6D, included within the derivation of the hourly rate for the Information Specialist, about 10 minutes for pre-session plus post-session activity. Even with allowances for inexperience or for testing activities, the 10 minute estimate is too low. A better estimate of steady-state, non-testing, activities occurring either before or after an appointment is probably 20 to 30 minutes. Thus, the hourly rate charged for a Specialist's time should be higher for operational cost recovery. In addition, if it becomes clear that there is a normal role for pre-appointment searches, then the cost for this computer activity will also need to be included within the pricing structure, either by adding that connect time to the appointment connect time, or by increasing the administrative surcharge within the computer search component of the pricing structure. We expect to look into this area more closely in Phase 2.

In addition, there are interactions which transpire between the Coordinator's Office and the Specialists. Each interaction is generally brief but the number is, in part, a function of the volume, and their aggregate represents Specialist time yet to be accounted for in the pricing structure for true cost recovery.

Another element of interaction is a weekly two-hour meeting of the Specialists with the Coordinator and with Laboratory staff to review the operational activities and associated problems, to discuss further development and testing and to continue training. In a steady-state operation, this activity should be less frequent. However, periodic review meetings between Specialists and Coordinator represent an activity that should also be accounted for in a revised pricing structure.

The initial pricing structure was also based on operations requiring a Coordinator estimated at 20 percent full-time and an Assistant to the Coordinator estimated at 60 percent full-time (cf.Task 6D). Our experience indicates that the duties performed by the Assistant reflect, in actuality, a full-time position. The Coordinator position is less clear but experience indicates that in a steady-state operation (with little or no development and experimentation) it is at least 70 percent, perhaps higher, for coordination-related duties. These duties include, but are not limited to: marketing and publicity; relations with users, Specialists, library staff and administration, NASIC at NEBHE, and the profession; review and dissemination of updated information to Specialists. The remainder of the Coordinator's time can be devoted to performing Information Specialist functions. Thus, the time assumptions underlying the administrative surcharge in the pricing structure need to be revised in light of our initial experience in order to recover costs.

Price stability is a matter of some concern. The advantages to planning and marketing of a relatively stable price schedule should be obvious. At M.I.T., the price structure reflects elements of expense. When the expense rates change, it is in keeping with cost recovery to eventually pass along these changes in the cost to the user. When to pass along the cost is another matter. For those expenses under M.I.T. control, such as administrative expenses or Specialist charges, a periodic review using historical data for the past period can be implemented. An annual review, perhaps coincident with the end of the academic year, is probably sufficient. Other expenses not under M.I.T. control, such as computer or communication expenses, present difficulties because (1) they represent the bulk of user charges, (2) they typically are subject to change on 30 day written notice, and (3) these expenses reflect a direct outflow of funds from M.I.T. (or any other institution). Because of the size and nature of this expense, it is one that should be passed on to users as of the date the new expense rate becomes effective. However, a short advance notice period from suppliers is insufficient time to prepare, produce, announce, and disseminate new price lists to users in advance of the change date. The effects of a short notice on marketing are unwelcome. In an academic market, a stable period of at least a semester is highly desirable although a full year is even more preferable. NASIC/Central at NEBHE, on behalf of the institutions it represents, needs to negotiate with systems for a period of advance notice of price changes that would encourage price stability within academic calendar time frames.

The remarks in the paragraph above on price stability are based upon experience with a month notice by SDC of a price change (it was their first change in over a year) for the connect hour rates of some data bases and on a change in the basis of printout costs (from a per-page cost to a per-citation cost). By way of contrast, Georgia issues a price list effective for a 12 month period and will honor current rates for specific continuing searches that extend beyond the price-year. On 1 February 1974 we did change our price rates to reflect the SDC changes. During the few weeks preceding the change, inquirers were verbally told about the forthcoming change. They were even told that if they made an appointment before 1 February, it would cost them less. There was no discernible effect of the price change on either data base usage or on printout requests. This is probably because we were still too new for many potential users to have been fully aware of the service, let alone the older rates.

The reader may be interested in the revenues that were generated in the two and a half month period since operations began. Table 9 summarizes the revenue by major categories for computer and administrative charges, Information Specialist charges, printout charges, and other charges. Associated with these categories are allowances against the charges to users for problems, technical and non-technical, that arose during an appointment. The largest allowance category is for the time of the Information Specialist since most users are eligible for the introductory credit. After all allowances, the total net revenue generated for this initial period is $1095.05 for NASIC services and $494.08 for MEDLINE services. In phase 2, we expect to be able to relate revenue to expenses.

Appendix B contains most of the forms that we have been using during our initial operations. These forms were described under Task 6F. A few of the forms have undergone modification since they were implemented and some still need to be revised but none of the changes are major. One initial form for an inquiry follow-up has since been eliminated because common techniques such as standard pads or quick notes or phone calls are more viable. The forms are used at different times by different people and for different functions. Sometimes the information entered onto the forms is less than complete but for the most part, everyone recognizes

Table 9

NASIC AT MIT

REVENUE SUMMARY DATA

15 NOVEMBER 1973 TO 28 FEBRUARY 1974

REVENUE SUMMARY:	NASIC		MEDLINE	
COMPUTER AND ADMINISTRATIVE CHARGES	$989.02		$388.30	
LESS: MACHINE CONNECTION PROBLEM ALLOWANCE:	(23.76)		(11.70)	
NET COMPUTER AND ADMINISTRATIVE CHARGES		$965.26		$376.60
INFORMATION SPECIALIST CHARGES:	$267.62		$261.14	
LESS: INTRODUCTORY CREDIT:	(220.55)		(239.56)	
LESS: OTHER SPECIALIST TIME ALLOWANCES:	(12.00)		(--)	
NET SPECIALIST TIME CHARGES:		$ 35.07		$ 21.58
PRINTOUT CHARGES:		$ 94.72		$ 98.70
OTHER CHARGES:	--		--	
LESS: OTHER ALLOWANCES:	(--)		(2.80)	
NET OTHER CHARGES:		--		($ 2.80)
TOTAL NET REVENUE		$1095.05		$494.08

the importance of the documentation and its role in characterizing, analyzing, and understanding the service operations and using this information as a basis for further experimentation, testing, or improvement.

When the problem statement that users are asked to fill out was drawn up, we had misgivings about its length. Our fear has proven to be unfounded. User receptivity of the statement has been gratifying. The statement has been particularly beneficial to the Specialist in being able to talk with the user in his own terms. The statement is also beneficial to the user in getting him to think more about his own problem and its boundary conditions in advance of the search so that the Specialist and user can optimize their interaction. In one instance, a potential user was attempting to provide a narrative of his problem but as a result he realized that he did not understand his problem. He stated that to go ahead with a search at that time would be a waste of his own time and money but that he would do a search after he thought more about what he was after. This phenomenon is not new to reference librarians but when users pay directly for services, then knowing what you are doing takes on added importance. Users are told that the efficiency of their appointment can be increased and that the cost of services to them can perhaps be lessened if the statement is filled out prior to their appointment. Users have filled out the statements in varying degrees of completeness depending upon just how much information they already have at their command. Only two users have raised any objection to the statement, and this only because of misunderstanding its purpose. If a user does not fill out the statement, and he is not required to, the necessary information is still gathered by the Specialist during a more extensive reference interview at the time of the appointment.

Interactive retrieval systems are dynamic and change with modifications over time. There are occasional operational difficulties, both technical hardware and logical software problems, and there are also data base content problems. Many problems make themselves known at the interface between searcher and system. Some difficulties are not so much a problem as they are a system procedure in need of either improvement, or further exposition. We have been in frequent contact with the SDC Search Service Staff in order to resolve whatever issues arise with respect to

ORBIT and the available data bases. It is a pleasure to note that the SDC staff has always been attentive, responsive, and most cordial to our inquiries. We have had occasion to offer a number of suggestions to SDC on improvements to their system. While we make no claim to either the originality or uniqueness of such suggestions we do note that SDC has acted on many of them. This is highly indicative of the kind of results that a central NASIC organization could help bring about even more effectively because it would represent a large user group. For items in need of clarification, a central NASIC organization could effectively disseminate responses and examples or solutions that are more extensive than the information typically appearing in retrieval system newsletters. We are already doing this kind of dissemination at M.I.T. from the Laboratory to the Coordinator's Office to the Specialists. There is a continuing need to generate materials that supplement the information available directly from a retrieval system.

The user receptivity of NASIC services is of considerable interest but we can give only an informal report at this time. In Phase 2, we will undertake a more formal survey of users (and non-users as well) to obtain personal reactions and to have a basis for changes to the operation. We have, however, received enough informal and unsolicited remarks by users to know that they do like and respond positively to the in depth, customized service and personal attention given to their bibliographic needs.

One of our earliest users ran three searches, one on CHEMCON and two on MEDLINE, to complete his background research. He said he was satisfied, that he had obtained very relevant information, and that peripheral material that at first did not seem pertinent had been found to be useful to other members of his research group.

A CHEMCON user, an older man and a faculty member, had been a frequent user of Chemical Abstracts in printed form. He related that he had been extremely pleased with his search results and he also said, "Frankly, I'm just getting too old to wade through all that stuff--I'd rather let the machine do it--it's really something." His particular search came to $57 and he said it was worth every cent.

Much of the user response is seen at the Coordinator's Office when people stop by to pick up their off-line printouts. They usually take a few minutes to review it and have universally seemed pleased with their results.

A graduate student from Harvard stopped by for the printouts from combined ERIC and MEDLINE searches on information gathering by eye movements. She said, "I could do this every day." The results were to be shared with her research group who were working on a proposal. Her search is also indicative of the multidisciplinary nature of many search problems. In this case, the topic was approached from both its physiological aspects and its psychological or reading aspects.

Our initial efforts in integrating computerized reference service with traditional service can also be illustrated. We saw earlier from Table 6 that a number of users have been referred to NASIC by the library reference staff. In turn, NASIC users are also referred to traditional sources by the Specialists who look to see how the user can best be served. A student who made an appointment to use INFORM opted for a traditional search after the Specialist told him of several printed indexes which were more relevant for his particular problem than the machine stored data base he had chosen. Another user, an administrative officer, who did search INFORM was also told by the Specialist of two printed materials of which he was previously unaware, one of the printed items being Business Periodicals Index. This individual not only also used the printed sources, but he has also returned to do a further search both by computer and in printed sources.

Thus, NASIC and traditional library services are complementing one another. The emphasis is on understanding and solving the user's problem using the most appropriate sources and thereby gaining the fullest utility of information resources regardless of format or media.

Analysis for Future NASIC Systems (Task 13)

M.I.T. staff have held a discussion with staff members from QEI Inc. who were performing a study of evaluative procedures that will be useful for NASIC. This discussion included consideration of criteria for computer systems. QEI recommendations may be found in a separate report submitted by QEI to NEBHE.

M.I.T., NERCOMP (New England Regional Computing Program), and NEBHE staff have met to discuss potential models of NASIC operations. In the area of NASIC services, these include a batch model, a batch model with remote job entry input and/or remote printer output, and an on-line model. Communications models between a NASIC Information Specialist and the user include personal contact, voice phone, Telex-like connections, mail, and lastly, direct communication between user and supplier with NASIC acting only as comptroller. The consensus at this time is that there should not be any dedicated NASIC computer, that NASIC should not build a network parallel to existing regional networks, but that NASIC should consider using existing computing facilities in the region , and that the ability to provide on-line services may be the primary future mode of operation. Model definition will be refined in Phase 2 as additional information about existing facilities in the region and on actual experience with the M.I.T. pilot operations is gathered.

Plans (Task 14)

The discussion under many of the preceding tasks has indicated areas in need of further study, development, or testing. A formal plan for work on NASIC at M.I.T. during Phase 2 has been submitted to the New England Board of Higher Education.

Appendix A. Project Personnel

Electronic Systems Laboratory

Professor J. Francis Reintjes
Mr. Alan R. Benenfeld
Mr. Richard S. Marcus
Mr. Jorge R. Peschiera

The MIT Libraries

Miss Natalie N. Nicholson
Ms. Marjorie Chryssostomidis
Mr. Edgar W. Davy
Ms. Margaret E. DePopolo
Mr. William J. Duggan
Mrs. Patrica T. Gordon
Ms. Irma Y. Johnson
Mr. James M. Kyed
Ms. Ann S. Longfellow
Ms. Margaret A. Otto
Ms. Mary E. Pensyl
Mr. Phillip W. Piper
Mr. Peter R. Scott
Mrs. Jacqueline Stymfal
Mrs. Frances B. B. Sumner
Ms. Nancy G. Vaupel
Ms. Susan E. Woodford

Information Processing Services

Mr. Robert H. Scott

Appendix B. Sample Forms Used in Operations

Copies of the following forms are included:

	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	ON-LINE DATA BASES
9-10:50 or 9:50-11	ANN SUSAN (MED)	X	PAT (med) X NANCY	ANN SUSAN (MED) JACKIE	JACKIE NANCY	BUSINESS — INFORM EDUCATION — ERIC MEDICINE — MEDLINE
11-12:30	SUSAN (med) *ANN	NANCY ANN	NANCY *PAT	JACKIE NANCY *ANN	SUSAN (med) *NANCY	CHEMISTRY — CHEMCON BUSINESS — INFORM MEDICINE — MEDLINE
12:30-2	PAT (med) MARJ	PAT (med) *SUSAN	SUSAN (med) MARJ	PAT (MED) MARJ	PAT (MED) ANN *MARJ	CHEMISTRY — CHEMCON BUSINESS — INFORM MEDICINE — MEDLINE
2-3:30	PAT (MED) JACKIE *MARJ	SUSAN (med) JACKIE *PAT	PAT (MED) JACKIE *ANN	X	PAT (med) MARJ *SUSAN	BUSINESS — INFORM EDUCATION — ERIC MEDICINE — MEDLINE
3:30-5	NANCY *JACKIE	SUSAN (MED) MARJ *JACKIE	SUSAN (MED) ANN	X		BUSINESS — INFORM EDUCATION — ERIC MEDICINE — MEDLINE

*OFFLINE APPTS. ONLY.

(MED) MEDLINE PREP.

(med) MEDLINE BECOMES 2ND CHOICE.

Fig. B1 Initial Service Schedule for Appointments

Page 1 of ____

N A S I C A T M I T

INQUIRY DATA - GENERAL

Receiver: ______________ In-Person Phone Mail Date: __________ Time: __________

Location: CO BE D H R S Other ______________ Duration: ______________

Repeat Caller: Y N Repeat User: Y N Agency Call: ______________

CALLER: Name: ______________________________

Address: ______________________________

____

Phone and Hours: ______________________________

Alternate Address/Phone: ______________________________

Dept. or Lab. or Course: ______________________________

Status: MIT Non-MIT Affil: ______________________________ (ILO P A)

Faculty DSR Staff Library Graduate Student (D M) (TA RA)

Post-Doc. Admin. Staff Undergraduate (1 2 3 4 5)

Other Position: ______________________________

Is Inquiry for Someone Else? ______________________________

FOR NCO USE ONLY: Appt. for: ______________________________

Appointment: Day: __________ Time: __________ Specialist: __________

Location: CO BE D H R S Other ______________________________

Special Set-Ups: ______________________________

Consultation to Discuss Services Available: General For Specific Problem

Expected Search Services: Retrospective On-Line Data Base(s) __________

SDI Off-Line System(s) __________

TBD TBD TBD

Other Services: ______________________________

____

Expected Payment Method: MIT Charge (Acct. ______) Check Cash Non-MIT P.O.

Other (including experimental): ______________________________

Brief Problem Title: ______________________________

____

____

11/73

Fig. B2 Inquiry Data - General

Page____ of______

N A S I C A T M I T

INQUIRY DATA - DESCRIPTION

Caller:________________________________ Date:__________________

Codes	Follow Up *	Description (Include summary, problems, actions required, comments)

1. General
2. Brochures/Publ.
3. Data Bases/Retr. Syst.
4. Searching
5. Pricing
6. Specialists
7. Terminals
8. Outputs
9. Output Recvd.
10. Delivery Services
11. Service Hours/Locations
12. About Appts.
13. User Funds
14. Requisitions
15. User Worksheets
16. Work Order
17. Order
18. Change Order
19. Cancel Service
20. Billing
21. Credits
22. Payments
23. Connection Problems
24. Searching Problems
25. Equipment Problems
26. Experiments
27. Comments
28. Complaints
29. Arrange Appt.
30. Change Appt.
31. Cancel Appt.
32. Appt. Reminder
33. Appt. Scheduling Problem
34. Schedule Change
50. User Topic
80. NEBHE
81. Outside Agency
99. Other

11/73

Fig. B3 Inquiry Data - Description

Page ____ of ____

N A S I C A T M I T

INQUIRY DATA - SPECIAL QUESTIONS

Caller: ______________________________ Date: ______________

1. Several methods of announcing NASIC have been used. How did you specifically learn of NASIC?

_____ Brochure	_____ Tech Talk	_____ Library Bulletin
_____ Letter	_____ Poster	_____ Saw a Site
_____ Colleague	_____ Meeting	_____ Saw a Session

Other ______________________________

2 - 4. If no appointment has been made, and if it is not obvious from previous data:

2. Can you tell us the reason you do not wish to arrange for an appointment at this time? Your answer may help us improve upon our services.

3. What is the subject area of interest to you?

4. Do you have access to an MIT charge account?
(If yes) Who would need to approve a requisition against that account?
(If no) Are other funding sources available to you?

11/73

Fig. B4 Inquiry Data - Special Questions

Room 10-400 | NASIC AT MIT | 253-7746

Request for Information

Brochures about NASIC search services are available at the reference desk at each MIT library. For further information about NASIC services, kindly phone 253-7746 or stop by the central NASIC Coordination Office, Room 10-400, between 9 and 5, Monday thru Friday. If you prefer, let us know when and where we may contact you. Please leave this card with a Library staff member or put it into the Institute mail. Thank you.

Mary Pensyl, NASIC Coordinator

Name:

Address:

Phone(s):

Hours You May be Reached:

Nature of Inquiry:

TO: MARY PENSYL
NASIC COORDINATION OFFICE
ROOM 10-400
M.I.T.

Fig. B5 Inquiry Post Card

MASSACHUSETTS INSTITUTE OF TECHNOLOGY CAMBRIDGE, MASSACHUSETTS 02139

Room 10-400 MIT LIBRARIES 253-7746

N A S I C A T M I T

APPOINTMENT REMINDER

To: Date:

You have an appointment

with

at

on

at

Please be prompt. If you are charging NASIC services to an MIT account be sure to bring an authorized requisition slip with you. If you must change your appointment, please call the NASIC Coordinator's Office, 253-7746.

You can help increase the efficiency of your appointment with the Information Specialist and perhaps lower the cost of services to you by carefully filling in the attached User Problem Statement <u>before</u> your appointment. An initial search strategy will be developed by the Specialist together with you and it will be based upon your replies. The initial strategy may be modified by you and the Specialist as search results are received and reviewed. Kindly present these forms to the Information Specialist at the start of your appointment.

Thank you.

Mary Pensyl

Mary Pensyl
NASIC Coordinator

Fig. B6 Appointment Reminder

Page 1

W.O.N.________

NASIC AT MIT

USER PROBLEM STATEMENT

Name:______________________________ Appt. Date: ______________

Completing this form prior to your appointment will increase the efficiency of NASIC service during your appointment and will likely lower the cost of service to you.

1. Please give in your own words a narrative description of the problem to be searched. Be specific. Define phrases with special meaning. Cover all aspects of the problem but please underline particular phrases that are more important to you. Append a list to your narrative of any synonyms, closely-related phrases, and alternate spellings. Please indicate if any words or phrases have a special use that you wish to exclude. Use scientific and technical as well as common vocabulary.

11/73

Fig. B7 User Problem Statement

User Problem Statement contd. Page 2.

Name:______________________________ W.O.N.__________

2. Unless already stated, please indicate any models, end uses, or applications of your problem that could be helpful in retrieving useful references.

____

____

____

3. Please state any topics related to (or applications of, or views of, or approaches to) your specific problem that are not of interest if you wish to exclude retrieving references to any documents on such topics.

____

____

____

4. Please give a title to your problem.

____

____

5. Please list two or three of the most important authors (and/or organizations) publishing on your topic. Complete names are helpful. Please indicate if you wish to exclude documents by any of these (or other) authors or organizations because of prior familiarity with their publications.

____

____

____

6. Please list two or three of the most important journals covering your problem. Please indicate if you wish to retrieve references to documents from only these journals. Please indicate if you wish not to retrieve references to documents from any particular journal, perhaps because you personally receive the journal.

____

____

____

7. Do you wish either to retrieve or not retrieve references to documents written in a particular language? _____Does not matter _____Retrieve English only
Retrieve only in______________ Do not retrieve in________________.

8. Do you wish to exclude references to particular types of documents?
Exclude _____Journal articles _____Books _____Patents _____Reports
_____Conference Papers _____Dissertations

11/73

Fig. B7 continued.

User Problem Statement contd. Page 3.

Name:________________________________ W.O.N.__________

9. Please list the complete citations to two or three of the most useful articles on your search topic. (It may be helpful to bring these articles with you to your appointment.)

____

____

____

____

____

____

10. Would you prefer

_____ a comprehensive search that retrieves most of the references relevant to your problem, but which may also retrieve many references not relevant to your problem?

_____ a narrow search that may retrieve fewer references relevant to your problem, but which also retrieves fewer non-relevant references?

11. Can you estimate the number of relevant documents

a) you think may be present in the literature _______

b) you would like to retrieve and get references for ______

12. If you have previously done a literature search (manually or by computer) on this problem or a closely related problem, please indicate if possible what was searched, what difficulties were encountered, and the overall result of the search.

____

____

____

____

____

____

____

1/73

Fig. B7 continued.

N A S I C A T M I T

WORK ORDER - PART 1 - SUMMARY

Specialist ______________ Service Date ________

Work Order No. ______________ NASIC Account ________

Name ______________

Address ______________

____

Phone ______________

Bill To (if different) ______________

☐ MIT Requisition No. ________ Dated ______ User Account No. ________.788 (NASIC)

☐ Purchase Order No. ________ Dated ______ ________.789 (MED)

☐ Cash Receipt No. ________ Dated ______ Amount Paid $ * ______ Account 11305.155

☐ Check ID ________ Amount Paid $______ Account 11305.155

MIT Personal Charge ☐ Employee ☐ Student MIT ID ________

SERVICES (*Minimum Charge applies): ▨ Industrial Rates Apply ☐	Service Rate	Units Used	Cost	Object Code
*Retrospective Searches (System, Data Base, Offline Coverage)				
*Partial Volume Retro. Search (System, Data Base, No. Issues)				
Current Awareness (System, Data Base, Issues)				
*Specialist Services (Description, Time) Specialist Appointment Time				159
Output Costs (Describe) Off-Line Printouts				161
*Document Services (Describe)				
Other Services (Describe)				
Consultation About Services (No Charge)				
	SUBTOTAL			—
Credits Deducted (Describe) Specialist Time			( )	163
Computer Problem Allowance			( )	173
Other Allowance			( )	179
	CREDITS SUBTOTAL			—
	TOTAL CHARGE			
	Charges Prepaid			
	Balance Due			

Supplementary Services to be Billed Later:

☐ Printout ☐ Delivery ☐ ________ Initials ____

☐ Prepayment of $________ exceeds Total Charge. REFUND of $__________ is due user

Thank you for using *NASIC AT MIT*. Call 253-7746 should you have further questions.

Fig. B8 Work Order

Page ____ of ____

N A S I C A T M I T

APPOINTMENT LOG AND REVIEW ANALYSIS

Specialist: ______________________ Date: ______________

User: ______________________ Work Order: ______________

Please record running notes of problems and important decisions during an appointment. Later, complete the notes with a more detailed commentary and analysis. In particular, note (1) technical problems with connections or terminals (e.g. nature, time, duration, attempts to solve); (2) search software problems (e.g. nature, solutions); (3) search strategy and performance problems (e.g. nature, development, effectiveness); (4) user interface behavior; (5) user commentary. For connection and software problems, attach if possible the relevant sections of printout.

Technical Problem		Notes, Descriptions, Commentary, Analyses
Time	Duration	

11/73

Fig. B9 Appointment Log and Review Analysis

Specialist ____________

Week Ending ____________

INFORMATION SPECIALIST TIME ALLOCATION SHEET

ACTIVITY	BASIC SERVICES							MEDLINE SERVICES						
	WKND	M	T	W	T	F	TOTAL	WKND	M	T	W	T	F	TOTAL
Training and Practice 1. Meetings														
2. Offline Search Preparation														
3. Online Search Sessions														
Operations and Services 4. Meetings														
5. Offline Preparation														
6. Online Preparation														
7. Appointments Not At Terminal														
8. Appointments At Terminal														
9. Delegated Online Searches														
10. Coord. Office Interactions														
General 11. Study														
12. Documentation														
13. Travel														
14. Other														
TOTAL EXPENDED TIME														
15. Extra-Curricular Time														
CHARGEABLE TOTAL														
Terminal Connect Time 16. Training and Practice														
17. Operations and Services														
CONNECT TIME TOTAL														

Fig. B10 Information Specialist Weekly Time Allocation Sheet

ON-LINE CONNECTION LOG

INFORMATION SPECIALIST____________________ MONTH__________197___

Date	User Name or Train or Demo	System	Data Base	Account (add holder if nec.)	Actual Connect Time (minutes)	Offline Citations	Ind. Rate	Connection Problem	Connect Allowance (minutes)

Fig. B11 Information Specialist On-Line Connection Log

CA/CONDENSATES

ACADEMIC RATES PER MINUTE

HOURLY RATE: $ 67

MIN	CHARGE	MIN	CHARGE
1	$ 1.12	51	$ 56.95
2	2.23	52	58.07
3	3.35	53	59.18
4	4.47	54	60.30
5	5.58	55	61.42
6	6.70	56	62.53
7	7.82	57	63.65
8	8.93	58	64.77
9	10.05	59	65.88
10	11.17	60	67.00
11	12.28	61	68.12
12	13.40	62	69.23
13	14.52	63	70.35
14	15.63	64	71.47
15	16.75	65	72.58
16	17.87	66	73.70
17	18.98	67	74.82
18	20.10	68	75.93
19	21.22	69	77.05
20	22.33	70	78.17
21	23.45	71	79.28
22	24.57	72	80.40
23	25.68	73	81.52
24	26.80	74	82.63
25	27.92	75	83.75
26	29.03	76	84.87
27	30.15	77	85.98
28	31.27	78	87.10
29	32.38	79	88.22
30	33.50	80	89.33
31	34.62	81	90.45
32	35.73	82	91.57
33	36.85	83	92.68
34	37.97	84	93.80
35	39.08	85	94.92
36	40.20	86	96.03
37	41.32	87	97.15
38	42.43	88	98.27
39	43.55	89	99.38
40	44.67	90	100.50
41	45.78	91	101.62
42	46.90	92	102.73
43	48.02	93	103.85
44	49.13	94	104.97
45	50.25	95	106.08
46	51.37	96	107.20
47	52.48	97	108.32
48	53.60	98	109.43
49	54.72	99	110.55
50	55.83	100	111.67

2/73

CA/CONDENSATES

COMMERCIAL RATES PER MINUTE

HOURLY RATE: $ 82

MIN	CHARGE	MIN	CHARGE
1	$ 1.37	51	$ 69.70
2	2.73	52	71.07
3	4.10	53	72.43
4	5.47	54	73.80
5	6.83	55	75.17
6	8.20	56	76.53
7	9.57	57	77.90
8	10.93	58	79.27
9	12.30	59	80.63
10	13.67	60	82.00
11	15.03	61	83.37
12	16.40	62	84.73
13	17.77	63	86.10
14	19.13	64	87.47
15	20.50	65	88.83
16	21.87	66	90.20
17	23.23	67	91.57
18	24.60	68	92.93
19	25.97	69	94.30
20	27.33	70	95.67
21	28.70	71	97.03
22	30.07	72	98.40
23	31.43	73	99.77
24	32.80	74	101.13
25	34.17	75	102.50
26	35.53	76	103.87
27	36.90	77	105.23
28	38.27	78	106.60
29	39.63	79	107.97
30	41.00	80	109.33
31	42.37	81	110.70
32	43.73	82	112.07
33	45.10	83	113.43
34	46.47	84	114.80
35	47.83	85	116.17
36	49.20	86	117.53
37	50.57	87	118.90
38	51.93	88	120.27
39	53.30	89	121.63
40	54.67	90	123.00
41	56.03	91	124.37
42	57.40	92	125.73
43	58.77	93	127.10
44	60.13	94	128.47
45	61.50	95	129.83
46	62.87	96	131.20
47	64.23	97	132.57
48	65.60	98	133.93
49	66.97	99	135.30
50	68.33	100	136.67

Fig. B12 Rate Sheet (illustrating MIT charge per minute for on-line connection using CA Condensates)

Appendix C. Sample Publicity Brochures

Copies of the following brochures are included:

Figure C1.	NASIC at MIT	- General Brochure
Figure C2.	NASIC/CA CONDENSATES	- Services for chemistry and chemical engineering
Figure C3.	NASIC/ERIC	- Services for education, linguistics and information science
Figure C4.	NASIC/INFORM	- Services for business management
Figure C5.	MEDLINE	

NASIC* AT MIT

Automated Bibliographical Services For Research

NEW MIT SERVICE

A new service of the MIT Libraries will be available under the auspices of NASIC, on a fee-for-service basis, at five divisional libraries:

Science — Dewey
Humanities — Rotch
Barker Engineering

Monday through Friday, on an appointment basis beginning November 15, 1973.

The NASIC service will open on an experimental basis and provide access to data bases in the following fields:

- Chemistry & Chemical Engineering
- Education, Linguistics, Information Sciences
- Business, Management, Economics, Marketing
- Medicine, Biology & Related Sciences

Both on-line and off-line access to the several data bases will be offered. An on-line search can produce a printed list of references that you can take with you. Full bibliographies can also be printed off-line and sent by mail. In some fields you can also be alerted to new publications as they appear, on a regular basis.

An Information Specialist will be available at each location to explain the system and to show you how to find recent publications relevant to your research interest.

For information about types of services available and associated costs, and to arrange for an appointment with an Information Specialist, contact the NASIC Coordinator's office:

253-7746

Room 10-400

*Northeast Academic Science Information Center

A Program of the New England Board of Higher Education, NASIC is supported by the National Science Foundation under Grant No. GN37296.

DATA BASES AVAILABLE

Data bases for the fields listed below are ready now.

Chemistry & Chemical Engineering

CHEMCON, for Chemical Abstracts Condensates, derives from Chemical Abstracts, sponsored by the American Chemical Society, and has the same coverage: about 6,000 articles selected from 10,000 journals are added each week. The on-line file goes back to 1970, the off-line file to 1968.

Education, Linguistics, Information Sciences

The ERIC (Educational Resources Information Center) data base is maintained by the U.S. Office of Education. Each month about 1,000 new reports and 1,500 new journal articles selected from over 500 journals are added to the on-line and off-line files, which go back to 1966.

Business, Management, Economics, Marketing

The INFORM data base produced by ABI, Inc., is updated monthly at a rate of about 900 articles selected from 280 journals for this on-line file, which goes back to August 1971.

Medicine, Biology & Related Sciences

MEDLINE, operated by the National Library of Medicine, indexes the 1,200 leading journals in the biomedical field since 1970. It covers about 60% of the material in Index Medicus, with about 12,000 articles being added each month to this on-line file.

The complete NASIC information service now being planned will eventually include data for all major fields of research interest at MIT. New data bases will be added within the next few months to cover interests in government research, engineering and physics. All these data bases will be available through the NASIC regional network except MEDLINE, which is available on-line by separate arrangement with the National Library of Medicine.

(1) (2)

Fig. C1 NASIC at MIT - General Brochure

KINDS OF SERVICE

Two broad classes of service are being offered: retrospective searching of data bases, and a current awareness (alerting) service. All data bases can be searched on-line, and either immediate on-line printouts or delayed off-line printouts are possible. Most data bases can also be searched in an off-line mode.

The on-line search mode enables the operator to converse directly with the computer and to obtain an immediate response to a query in the form of a printed list of citations. This interactive feature makes it possible to use the system in an exploratory way to improve the effectiveness of the search. The operator can modify the search words and adjust the strategy as the search progresses in order to achieve a closer match with the needs of the researcher. This exploratory capability with the aim of refining the definition of the bibliographic problem is one of the most important features of the system.

If the list of citations is long and the need is not immediate an additional option permits printing off-line and delivery by mail, at substantial savings.

Also available for some data bases is an off-line retrospective search service. This option may result in lower cost for extensive search and printout requirements.

A current awareness service option that will alert you at regular intervals to new publications in your field is available for several of the data bases.

Search magnitude can be limited in various ways to a partial data base and by time, author, institution and other requirements.

USER ASSISTANCE

The assistance provided by the Information Specialist is an essential part of the new service. Each of the information retrieval systems has its peculiar special language, logic, rules of access, procedures, policies, kind of information and form of printout. The Information Specialist is familiar with all of the available data bases. This knowledge will be particularly valuable in formulating search strategies, interacting with the system and dealing with interdisciplinary information requests.

The primary task of the Information Specialist is to assist you in translating your problem statement into the language of the system in order to help you to maximize the satisfaction you derive from the system and to minimize the cost of making a search. This user interaction may take half an hour or more to develop an appropriate search strategy.

(3)

SPONSORSHIP

The NASIC computer-based bibliographic service is being developed by the New England Board of Higher Education under a grant from the National Science Foundation. The experimental MIT service is being tested by the MIT Libraries and Electronic Systems Laboratory under a contract with NASIC. The experimental service at MIT will become the first node of a regional network of science information centers located at university libraries in the northeast region. Policies and procedures for the NASIC network will be based on experience gained at MIT during this experimental period.

The MEDLINE system for the biomedical sciences is not a part of the NASIC service, but is made available at the same terminals at MIT by arrangement with the National Library of Medicine.

COST

Although development costs are being underwritten by NSF, operating costs must be recovered on a fee-for-service basis. Fees vary with the amount of service provided by the Information Specialist, the data base searched and the time spent at the terminal.

Since MEDLINE is substantially subsidized by the National Library of Medicine, the cost of searching this data base is less than search costs for the others.

The price structure includes a fee for the Specialist's time spent in developing search strategies with you and for operating the system. Because time spent at the terminal is expensive, users can generally minimize overall costs by taking advantage of the skills of the Information Specialist. There is a charge of $6.00/hour of Specialist's time with a minimum charge of $5.00.(A)

Further details of fees are included in separate brochures that describe each data base. Typical examples of computer search costs are: for a half hour of time spent at the terminal searching the business data base INFORM -- $33.50, for a similar search of MEDLINE -- $5.00, for ERIC -- $22.00, for CHEMCON -- $27.50. Times and prices may be lower for simple problems or higher for more complex problems or those that have not been well defined initially by the user. An off-line search of a one year collection of ERIC would be $76.00. A Current Awareness subscription service for the Chemistry data base would cost approximately $7.50 for each week the service is rendered. Off-line printing charges are at the rate of 10¢ per printed page.

For further information and to arrange for an appointment with an Information Specialist, call: 253-7746.

(A)During the initial "break-in" period, users will receive a credit for the Information Specialist's time up to a maximum credit of $50.00. This credit is good until the end of the academic year, in June 1974.

(4)

Fig. C1 continued.

NASIC*/CA-CONDENSATES

AUTOMATED BIBLIOGRAPHIC SERVICES FOR RESEARCH

SERVICES FOR CHEMISTRY AND CHEMICAL ENGINEERING

THE DATA BASE

CA-Condensates is the chemistry data base corresponding to the publication Chemical Abstracts produced by the Chemical Abstracts Service of the American Chemical Society. The worldwide data gathering capability of CAS provides comprehensive coverage of the literature in all fields of chemistry and disseminates bibliographic information in this literature in both printed and machine-readable form. The data base is issued on a weekly basis, each issue covering one half of the total subject scope of the data base. Searches may be tailored to the odd or even numbered issues.

NASIC/CA-CONDENSATES AT MIT

Computer-based bibliographic services including CA-Condensates add a new dimension to information retrieval traditionally performed by manual techniques. NASIC services, available through the MIT Libraries, enable you to employ a more exhaustive combination of retrieval parameters at relatively low cost to produce rapid and highly relevant search results.

USER ASSISTANCE

The CA-Condensates data base is accessible in interactive on-line or remote batch modes. MIT Information Specialists are available to assist you in the use of these services. For information about types of services available and associated costs, and to arrange for an appointment with an Information Specialist, contact the NASIC Coordinator's office:

253-7746

ROOM 10-400

*NORTHEAST ACADEMIC SCIENCE INFORMATION CENTER

A Program of the New England Board of Higher Education. NASIC is supported by the National Science Foundation under Grant No. GN37296.

(1)

NASIC/CA-CONDENSATES

SUBJECT AREAS

Chemistry and Chemical Engineering related topics are covered as in Chemical Abstracts, in five major sections:

1. Biochemistry Sections (CBAS)
2. Organic Chemistry Sections (CAOS)
3. Macromolecular Sections (CAMS)
4. Applied Chemistry and Chemical Engineering Sections (CAAS)
5. Physical and Analytical Chemistry Sections (CAPS)

COVERAGE

The CA-Condensates data base covers the chemistry-related literature published in over 12,000 journals as well as patents issued by 26 countries. New books, conference proceedings, and government research reports are regularly monitored to select those documents pertinent to the chemical sciences. The on-line data base references information from the several issues of Chemical Abstracts published since 1970. The off-line data base, begins with Chemical Abstracts Volume No. 69, first published in July 1968.

FILE SIZE AND UPDATING

The on-line CA-Condensates data base presently contains records for over 1,100,000 documents, while the off-line data base contains over 1,500,000 records. Approximately 14,000 new records are added to the data base each month. The on-line file is updated biweekly. The off-line file is maintained in two parts: biochemistry and organic chemistry in one part (corresponding to the odd-numbered issues of Chemical Abstracts) and the other three sections in a second part (even numbered issues of Chemical Abstracts). Each part is updated separately on an alternating week basis.

RECORD CONTENT

The CA-Condensates data base includes the following information elements from the corresponding issues of Chemical Abstracts: titles of papers, patents, reports; names and organizational affiliation of authors and/or assignees; bibliographic citations, language of document and subject indexing.

(2)

Fig. C2 NASIC/CA-CONDENSATES Brochure

NASIC/CA-CONDENSATES

ACCESS OPTIONS

NASIC's computer-based bibliographic data bases on Chemistry and Chemical Engineering literature are available for search in on-line or off-line modes.

SERVICES AVAILABLE

Current awareness and retrospective search services tailored to your specific interests, are now available at the MIT Libraries. The Current Awareness Service provides routine periodic notification of the most recent publications which match the subscribing researcher's interest profile. Retrospective Search Services, generally covering several years of publications, are also available on-line or off-line. For many of the citations obtained through your NASIC search, you may obtain through the MIT Libraries a photocopy (or in some cases, hard copy or a microfilm copy) of the full text of the document.

COST OF SERVICES

Charges to academic users will be based on the following rates:

Current Awareness $370 (annual subscription, weekly delivery) (A)

Retrospective Search

On-line: $55 per connect hour at terminal; minimum charge $5.

Off-line: $166 per year of data base searched(A)

Information Specialist assistance; $8 per hour; minimum charge $5.(B)

Off-line computer printouts @ ten cents per page (4x6 card stock also available @ two cents each extra)

(A)Half-price if only odd or even issues searched

(B)During the initial "break-in" period, users will receive a credit for the Information Specialist's time up to a maximum credit of $50.00. This offer expires June 1974.

NASIC - A REGIONAL RESOURCE

NASIC - The Northeast Academic Science Information Center - is being developed by the New England Board of Higher Education (NEBHE) to provide the Northeast area with a central access point to the nation's growing and diverse information resources in computer-readable form. This development is being aided by staff of the Association of Research Libraries, the Massachusetts Institute of Technology and by other organizations and consultants working under subcontract to NEBHE.

By aggregating data bases and existing information services, NASIC provides comprehensive and in-depth services to users. NASIC thus aids in increasing the capabilities of the Northeast's academic community.

The increasing availability of computer-readable data bases now makes it possible for R&D personnel to keep up with the proliferation of professional journals and with the growing record of experimental and statistical data. Computers permit searching of hundreds of thousands of references in the time it would take a human researcher to read one page.

NASIC AT MIT

To assist in meeting the information needs of the MIT community, a number of computerized bibliographic services are already available for several subject disciplines. Others will soon be added and, eventually, all major fields of research interest will be covered.

For further information on all computer-based services available at the MIT Libraries, contact the NASIC Coordinator's office:

253-7746

Room 10-400

THE NEW ENGLAND BOARD OF HIGHER EDUCATION
40 Grove Street
Wellesley, Massachusetts 02181
(617) 235-8071

11/73 - 1M

(3) (4)

Fig. C2 continued.

NASIC*/ERIC

AUTOMATED BIBLIOGRAPHIC

SERVICES FOR RESEARCH

SERVICES FOR EDUCATION, LINGUISTICS

AND INFORMATION SCIENCE

THE DATA BASE

ERIC (Educational Resources Information Center) is the educational data base developed and maintained by the U.S. Office of Education. Eighteen clearinghouses located throughout the United States, and now reporting to the National Institute of Education, collect, screen, index, and abstract the report and periodical literature in education and education-related fields.

NASIC/ERIC AT MIT

Computer-based bibliographic services including ERIC add a new dimension to information retrieval traditionally performed by manual techniques. NASIC services, available through the MIT Libraries, enable you to employ an exhaustive combination of retrieval parameters at relatively low cost to produce rapid and highly relevant search results.

USER ASSISTANCE

The ERIC data base is accessible in interactive on-line or remote batch modes. MIT Information Specialists are available to assist you in the use of these services. For information about types of services available and associated costs, and to arrange for an appointment with an Information Specialist, contact the NASIC Coordinator's office:

253-7746

Room 10-400

*NORTHEAST ACADEMIC SCIENCE

INFORMATION CENTER

A Program of the New England Board of Higher Education, NASIC is supported by the National Science Foundation under Grant No. GN37296.

NASIC/ERIC

SUBJECT AREAS

Education and education-related topics in ERIC include:

- Adult Education
- Counseling & Personnel Services
- Disadvantaged
- Early Childhood Education
- Educational Management
- Educational Media & Technology
- Exceptional Children
- Higher Education
- Junior Colleges
- Languages & Linguistics
- Library & Information Sciences
- Reading & Communication Skills
- Rural Education & Small Schools
- Science, Mathematics & Environmental Education
- Social Studies/Social Science Education
- Teacher Education
- Tests, Measurement & Evaluation
- Vocational & Technical Education

COVERAGE

The ERIC data base covers educational literature published since 1969* and contains all citations published in Research in Education (RIE) and Current Index to Journals in Education (CIJE), the two major printed monthly products of the ERIC system.

FILE SIZE AND UPDATING

The ERIC file currently contains records for over 135,000 documents. Approximately 1000 new reports and 1500 new journal articles selected from over 500 journals are added monthly into the ERIC file.

RECORD CONTENT

The ERIC record includes the following information for each document: the title, author name(s) and organizational affiliation, the publication citation (when and where published), and availability data (including price for microfiche or paper copy from ERIC), subject indexing and sponsoring agency with contract or grant number. RIE also has an abstract for all primary documents. Searching is possible on any item of information in the record.

*A limited number of documents going back to 1956 is also included.

(1) (2)

Fig. C3 NASIC/ERIC Brochure

NASIC/ERIC

ACCESS OPTIONS

NASIC's computer based bibliographic data bases on Educational Literature are available for search in on-line and off-line modes.

SERVICES AVAILABLE

Current awareness, and retrospective search services tailored to your specific interests, are now available at the MIT Libraries. The Current Awareness Service provides routine periodic notification of the most recent publications which match the subscribing researcher's interest profile. Retrospective Search Services, generally covering several years of publications, are also available on-line and off-line. For many of the citations obtained through your NASIC search, you may obtain through the MIT Libraries a photocopy (or in some cases, hard copy or microfilm copy) of the full text of the document.

COST OF SERVICES

Charges to academic users will be based on the following rates:

Current Awareness: $85(A) (annual subscription, quarterly delivery)

Retrospective Search

On-line: $44 per connect hour at terminal; minimum charge $5.

Off-line: $76 per year of data base searched (A,B)

Information Specialist assistance: $8 per hour; minimum charge $5.(C)

Off-line computer printouts @ ten cents per page (4x6 card stock also available @ two cents each extra)

(A) Half charge if only RIE or CIJE searched

(B) RIE search for 1956-68 is treated as one year @ $38.00

(C) During the initial "break-in" period, users will receive a credit for the Information Specialist's time up to a maximum credit of $50.00. This offer expires June 1974.

(3)

NASIC - A REGIONAL RESOURCE

NASIC - The Northeast Academic Science Information Center - is being developed by the New England Board of Higher Education (NEBHE) to provide the Northeast area with a central access point to the nation's growing and diverse information resources in computer-readable form. This development is being aided by staff of the Association of Research Libraries, the Massachusetts Institute of Technology and by other organizations and consultants working under subcontract to NEBHE.

By aggregating data bases and existing information services, NASIC provides comprehensive and in-depth services to users. NASIC thus aids in increasing the capabilities of the Northeast's academic community.

The increasing availability of computer-readable data bases now makes it possible for R&D personnel to keep up with the proliferation of professional journals and with the growing record of experimental and statistical data. Computers permit searching of hundreds of thousands of references in the time it would take a human researcher to read one page.

NASIC AT MIT

To assist in meeting the information needs of the MIT community, a number of computerized bibliographic services are already available for several subject disciplines. Others will soon be added and, eventually, all major fields of research interest will be covered.

For further information on all computer-based services available at the MIT Libraries, contact the NASIC Coordinator's office:

253-7746

Room 10-400

THE NEW ENGLAND BOARD OF HIGHER EDUCATION
40 Grove Street
Wellesley, Massachusetts 021[illegible]1
(617) 235-8071

11/73 - 1M

(4)

Fig. C3 continued.

NASIC*/INFORM

AUTOMATED BIBLIOGRAPHIC SERVICES FOR RESEARCH

SERVICES FOR BUSINESS MANAGEMENT

THE DATA BASE

INFORM is a business management oriented data base produced by Abstracted Business Information, Inc. This data base covers approximately 280 journals specializing in the areas of finance, management, economics, statistics, business law and marketing.

NASIC/INFORM AT MIT

Computer-based bibliographic services including INFORM add a new dimension to information retrieval traditionally performed by manual techniques. NASIC services, available through the MIT Libraries, enable you to employ a more exhaustive combination of retrieval parameters at relatively low cost to produce <u>rapid</u> and <u>highly relevant search results</u>.

USER ASSISTANCE

The INFORM data base is accessible in the interactive on-line mode. MIT Information Specialists are available to assist you in the use of these services. For information about types of services available and associated costs, and to arrange for an appointment with an Information Specialist, contact the NASIC Coordinator's office:

253 7746

Room 10-400

*NORTHEAST ACADEMIC SCIENCE INFORMATION CENTER

A Program of the New England Board of Higher Education, NASIC is supported by the National Science Foundation under Grant No. GN37296.

(1)

NASIC/INFORM

SUBJECT AREAS

The INFORM data base reflects comprehensive coverage of the business literature through references to feature articles from well known journals including:

Banking
Bests Review H/L
Bests Review P/L
Business Horizons
California Management Review
Duns Review
Fortune
Harvard Business Review
Personnel
Personnel Journal
Public Utilities Fortnightly
Sales Management
Sloan Management Review
Technology Review (MIT)
Journal of Marketing
Journal of Taxation
Nations Business

COVERAGE

The INFORM data base covers the major business management-related literature published in over 280 journals since August 1971.

FILE SIZE AND UPDATING

The INFORM file contains approximately 10,000 records. An average of 900 records are now added monthly by ABI into the INFORM file. Updates are searchable independently to provide a current awareness service.

RECORD CONTENT

The INFORM record contains the title, author, abstract, the publication citation and subject indexing as well as other categories of information.

(2)

Fig. C4 NASIC/INFORM Brochure

NASIC/INFORM

ACCESS OPTIONS

NASIC's computer based bibliographic data base on business related literature is available for search in an on-line mode.

SERVICES AVAILABLE

Current awareness and retrospective search services tailored to your specific interests, are now available at the MIT Libraries. The Current Awareness Service provides periodic notification of the most recent publications which match the subscribing researcher's interest profile. Retrospective Search Services, generally covering several years of publications, are also available on-line. For many of the citations obtained through your NASIC search, you may obtain through the MIT Libraries a photocopy (or in some cases, hard copy or microfilm copy) of the full text of the document.

COST OF SERVICES

Charges to academic users will be based on the following rates:

Current Awareness (available through periodic on-line searching of the updates)

Retrospective Search

- On-line: $67 per connect hour at terminal; minimum charge $5.

Information Specialist assistance: $8 per hour; minimum charge $2.(A)

Off-line computer printouts @ ten cents per page.

(A)During the initial "break-in" period, users will receive a credit for the Information Specialist's time up to a maximum credit of $50.00. This offer expires June 1974.

NASIC - A REGIONAL RESOURCE

NASIC - The Northeast Academic Science Information Center - is being developed by the New England Board of Higher Education (NEBHE) to provide the Northeast area with a central access point to the nation's growing and diverse information resources in computer-readable form. This development is being aided by staff of the Association of Research Libraries, the Massachusetts Institute of Technology and by other organizations and consultants working under subcontract to NEBHE.

By aggregating data bases and existing information services, NASIC provides comprehensive and in-depth services to users. NASIC thus aids in increasing the capabilities of the Northeast's academic community.

The increasing availability of computer-readable data bases now makes it possible for R&D personnel to keep up with the proliferation of professional journals and with the growing record of experimental and statistical data. Computers permit searching of hundreds of thousands of references in the time it would take a human researcher to read one page.

NASIC AT MIT

To assist in meeting the information needs of the MIT community, a number of computerized bibliographic services are already available for several subject disciplines. Others will soon be added and, eventually, all major fields of research interest will be covered.

For further information on all computer-based services available at the MIT Libraries, contact the NASIC Coordinator's office:

253-7746

Room 10-400

THE NEW ENGLAND BOARD OF HIGHER EDUCATION
40 Grove Street
Wellesley, Massachusetts 02181
(617) 235-8071

11/73 - 1M

(3) (4)

Fig. C4 continued.

MEDLINE
at
MIT

MEDLINE

The MIT Libraries offer a new service:

Automated Search of the recent literature

of biology, medicine, and related sciences

through MEDLINE, an on-line, computer-stored bibliographical information service operated by the National Library of Medicine.

MEDLINE indexes the 1200 leading journals in the biomedical fields, and is more up-to-date than the published *Index Medicus*.

MEDLINE will be available to the MIT community, on a fee-for-service basis,

Monday through Friday, by appointment

at five divisional libraries:

- Science
- Barker Engineering
- Dewey
- Rotch
- Humanities

An Information Specialist will be on hand at these libraries to explain and operate the system, and to show you how to devise a search strategy that will identify recent publications relevant to your research interests.

An on-line search can produce a printed list of references that you can take with you. Full bibliographies can also be printed off-line and sent to you by mail. Special searches of the most recent citations to be included in a forthcoming issue of *Index Medicus* are also available.

For information about services and costs, and to arrange an appointment with an Information Specialist, call the MEDLINE Coordinator's Office: 253-7746, Room 10-400.

(1)

The data base

MEDLINE is an on-line, interactive bibliographical information retrieval system operated by the National Library of Medicine. The data base is stored in computers in different parts of the country, and is now accessible through an international communications network known as TYMSHARE.

Searching with MEDLINE gives faster and more up-to-date results. It indexes the 1200 leading biomedical journals, using the standard Medical Subject Headings, which are arranged in categories such as:

- Anatomical Terms
- Organisms
- Diseases
- Chemicals and Drugs
- Psychiatry and Psychology
- Biological Sciences
- Physical Sciences
- Health Care
- Biochemistry

MEDLINE includes about 60 percent of the material in *Index Medicus*,* It covers the last three years and is updated monthly. New citations are available several weeks before they can appear in the printed index.

The system now includes about 500,000 records, each of which contains these items:

- Author
- Title
- Journal citation
- Year
- Language
- Subject headings

The MEDLINE system is more versatile than the ordinary printed index because it may be searched not only by subject and author, but in several other ways. Subject headings and search results may be combined in various ways to achieve a close match with your research interests.

---

*Note, however, that full *Index Medicus* coverage is available with the SDILINE file described on opposite page.

(2)

Fig. C5 MEDLINE Brochure

Kinds of Service

MEDLINE is normally used for an on-line search of the complete data base, with an immediate printout of a list of all documents for the last three years that match the user's request.

A printout can present any combination of the bibliographical items included in the records. It can be a list of titles only, or of authors and titles, or it can include all the information stored. If a list is long and not needed immediately, it can be printed off-line at reduced costs and sent by mail.

The on-line, interactive feature of MEDLINE means that the user is in continuous conversation with the computer so that he can modify and refine his search as he goes along. A skillful operator can use the system in an exploratory way to improve the effectiveness of the search. A first attempt often yields a list too long or too short to be useful. The Information Specialist can suggest various techniques for broadening or narrowing the search, and various ways of combining lists to identify the relevant documents. This kind of exploratory work with the aim of refining the definition of the bibliographical problem is one of the most important uses of the system.

The primary task of the Information Specialist is to assist you in translating your problem statement into the language of the system in order to help you to maximize the satisfaction you derive from the system and to minimize the cost of making a search. This user interaction with the Information Specialist may take half an hour or more to develop an appropriate search strategy.

A subsystem known as SDILINE contains the citations to be included in the forthcoming issue of Index Medicus and is separately searchable. SDILINE differs from the main MEDLINE data base in several ways: it covers the full range of journals in Index Medicus rather than the MEDLINE selection, and each significant word in the title is searchable. (In MEDLINE, titles are not directly searchable).

Cost

The charge for assistance by the Information Specialist is $8 per hour with a minimum charge of $5.*

In addition, the charge for time spent using the terminal is $18 per hour with a minimum charge of $5. One-third of this fee goes to the National Library of Medicine which is subsidizing a major portion of the MEDLINE costs. The remainder goes to M.I.T. to recover its costs.

There is also a charge for off-line printouts at the rate of ten cents per printed page.

A typical search might take a half hout at the terminal and a total of one hour with the Information Specialist for a total charge of $17.*

Other Data Bases

The MEDLINE service is one element of a comprehensive program of bibliographical information-retrieval services now being planned by the MIT Libraries, to cover the major fields of research interest at MIT. The program is designed as an integrated service with a number of different data bases all available from the same terminal under the guidance of an experienced Information Specialist.

The integrated MIT service is currently an experimental node in a regional network of information centers in university libraries. This network is known as the Northeast Academic Science Information Center (NASIC), and is being developed by the New England Board of Higher Education under a grant from the National Science Foundation, with the cooperation of the MIT Electronic Systems Laboratory.

At present, MEDLINE is not a part of the NASIC system, but is made available at the same terminals at MIT through the cooperation of the New England Regional Medical Library Service (NERMLS).

*During the initial "break-in" period, academic users will receive a credit for the Information Specialist's time up to a maximum credit of $50.00. This credit expires in June 1974.

11/73

(3) (4)

Fig. C5 continued.

EFFECTIVENESS AND COST-EFFECTIVENESS CONSIDERATIONS FOR NASIC INFORMATION SERVICES OPERATION

by

J. W. Kuipers, F. W. Lancaster and R. W. Thorpe

Q.E.I. Report Prepared For The
New England Board Of Higher Education

October 20, 1973

Q.E.I., Incorporated
119 The Great Road
Bedford, Mass. 01730

EFFECTIVENESS AND COST-EFFECTIVENESS CONSIDERATIONS FOR NASIC INFORMATION SERVICES OPERATION

Table of Contents

TABLES

FIGURES

SUMMARY

This report gives the results of a study conducted by QEI, Inc. for the management of the program for the Northeast Academic Science Information Center (NASIC) which has been initiated by the New England Board of Higher Education. The NASIC program is supported by the National Science Foundation and is intended to serve the science community of the ten-state area of the northeastern United States. The program was initiated early in 1973.

The purpose of this study is to provide to the NASIC management such background information, together with appropriate discussion of options and alternatives, which could assist it in making decisions regarding the kind of information services to be provided by NASIC and methods to be used to provide these services. A particular emphasis in the study is a review of guidelines, acceptance criteria and methods for evaluation as they have a bearing on proposed NASIC services and products.

The thirteen sections of this report and the attachments are listed in the Table of Contents. Each section discusses a particular aspect of NASIC information services operations. A brief summary for each section is given below.

A. NASIC Objectives and Some Factors Influencing the Attainment of these Objectives (pp. 14-18)

The function of NASIC is to improve accessibility of information needed by the science community in the Northeast United States. The objective includes two aspects:

1. Improving the accessibility of information: searching of accumulated files of data or references in response to specific information needs or requests: retrospective searching.
2. Improving the exposure to science information: providing service to inform users on a continuing basis, of newly available data or information: current awareness service.

In these activities, NASIC has a role essentially of that of an interface between the scientific community and the universe of available bibliographic and data resources.

In terms of effectiveness, the goal of NASIC is to provide the maximum level of exposure/accessibility to relevant data bases for the science community. In terms of cost-effectiveness the NASIC goal is to achieve maximum exposure/accessibility per dollar expended.

This section discusses three broad factors that will exert a major influence on the performance of NASIC:

What data bases NASIC makes available.

How these data bases are acquired.

How the data bases acquired by NASIC are made available to potential users.

B. Factors Affecting the Performance of Information Retrieval and Dissemination Systems (pp. 19-34)

This section reviews the factors affecting system performance relating to two types of information retrieval operation: the delegated search and the non-delegated search. The major factors which influence the success or failure of the delegated search are shown in Figure 2:

Coverage and appropriateness of the system.

User's interpretation of system capabilities and limitations.

Mode of interaction with the system.

User's ability to describe his needs.

System assistance and guidance in formulation of request. (request forms, interviews, interactive procedure).

Searcher's interpretation of user requirements.

Ability of the vocabulary to describe concepts occurring in requests. Help given to searcher by vocabulary structure.

Searcher's own ability to construct logically sound and complete strategy.

Capabilities of searching software.

Indexing policy (e.g., exhaustivity level). Indexing quality and accuracy.

Searcher's interpretation of user requirements. Quality of the user's request statement. Quality of the document surrogate.

Major factors influencing the performance of a non-delegated search are also covered in this section and are given in Figure 3.

Knowing what these factors are assists in making decisions relating to proposed NASIC services and products.

C. Criteria Relating to the Selection of Data Bases (pp. 35-54)

A major emphasis in the NASIC program will be the provision of services based upon machine-readable data bases. The number of data bases available has increased dramatically in the last few years: 269 machine-readable data base systems were identified in a recent survey.

This section reviews the question of how NASIC can determine which data bases to select and with what order of priority. Criteria affecting the selection of data bases are discussed in detail and are summarized in Table 1. Criteria are discussed under the following headings:

Subject Matter of Data Bases
Cost Factors
Quality Considerations
  Coverage
  Time Factors
  Indexing and Vocabulary Factors
Implementation Factors

D. Acquisition of Data Bases or Data Base Services: Alternatives (pp. 55-65)

NASIC may offer services based on acquiring a data base for in-house use or on acquiring services from that data base from a service center already in existence.

Data bases are available on a straight purchase basis, although the majority must be leased or licensed. Generally a leasing arrangement allows the recipient organization to offer services for its own staff purposes only. If services are to be offered to a wider community on a fee basis, a licensing agreement must be made with the data base supplier.

Various aspects involved in the acquisition of bibliographic services from existing suppliers are reviewed in this section. NASIC will be faced with various service options and these are discussed. Where service from a particular data base is already available from a service center this mode of access will be the preferred one. Where a choice exists between purchase of service from another service center or from the producer of the data base, specific cost comparisons must be made.

E. Guidelines on the Choice of Service Centers (pp. 66-91)

With a number of different suppliers of services available, NASIC must decide which information service centers to use. In this section various considerations relating to the choice of service center are discussed.

The comparison or evaluation of service centers is closely related to the matter of what criteria users of information services apply for evaluation purposes. Table 4 includes a summary of some of these criteria and a discussion of these is covered in pp. 66-91 of this section. The principal criteria include: cost, direct charges and effort involved in use; response time; quality considerations; coverage, completeness recall, precision, novelty, accuracy of data; cost-effectiveness: cost per reference supplied, cost per new relevant reference supplied.

A major decision NASIC must face for several data bases is whether access should be made through a center offering off-line batch services or through a service center making the data base available for direct on-line interrogation. The various aspects of on-line services and alternatives are reviewed in this section in pp. 71-82. Table 5 includes a typical schedule of commercially available on-line service. Table 6 through 11 include summary data about the characteristics of current awareness system features and search and output characteristics which must be taken into account in determining what services will be of interest for NASIC operations.

To reinforce and supplement the material of this section a comprehensive report on "The Present Status of On-Line Interactive Retrieval Systems" by F. W. Lancaster, is provided as an attachment in pp. 158-206 of this report.

F. Some Considerations Relating to NASIC Products and Services (pp. 92-100)

This section reviews certain kinds of information services which NASIC will wish to consider for its operation. Information services are of two broad types: Current awareness services (such as selective dissemination: SDI) and Retrospective search services (on-demand).

It seems likely that NASIC should give first priority to the SDI aspects of its services. For SDI services the training of NASIC information services librarians will be of high importance. On-line SDI, group SDI, service to provide information on current research in specific disciplines are topics covered in pp. 93-95 of this section.

In retrospective searching quick-reference searches and comprehensive searches can be offered. An important tool to be generated for search services of this kind will be a printed guide to available resources.

Discussion regarding an interactive on-line capability at NASIC headquarters that is open to members of the scientific community, a referral service, support for the creation and exploitation of personal files, are topics covered in pp. 98-100 of this section.

G. Software Aspects of Information Service Center Operation (pp. 101-117)

Operations and procedures necessary to provide information services desired of NASIC will be largely computer-based. General criteria for selecting and evaluating computer software are presented in the section in order to assist NASIC in choosing among possible alternatives. The principal computer program areas of interest to NASIC are: input/output routines, search routines, data management routines, database transformation routines. In Table 13 evaluation and selection criteria for information retrieval software are summarized. These criteria include: documentation, reliability, throughput time, availability of updating, cost, maintenance, operating time, ease of usage, uniqueness, etc.

H. Software Considerations: Data Base Formats (pp. 118-125)

It is desirable for an information service center to convert all of its data bases into the same form and format in order to minimize searching software and processing. This topic together with discussion of various record formats and the use of file inversion to speed up processing, are covered in this section of the report.

I. Software Considerations: Search Operations (pp. 126-131)

The difference in search operations for current awareness services and retrospective searching are reviewed in this section. Particular emphasis is placed on the advantages and disadvantages of on-line or off-line operations for each type of search.

SUMMARY Cont'd.

J. Software Considerations: Input/Output Processes, Data Management Requirements (pp. 132-136)

This section discusses input/output routines and some data management requirements of a typical information services center.

Input routines such as syntactic analysis of English input and profile inversion are discussed. The advantages of "table-driven" output routines are mentioned and the data management aspects of service center operation concerned with such items as billing, accounting and privacy are included in p. 134 of this section. Because of the large size of data bases, the advantages and problems associated with means for speeding up processing, such as grouping of requests, is discussed.

K. Communications Aspects of Information Service Center Operations (pp. 137-140)

The successful operation of an information services center will be greatly dependent upon the nature of its supporting communications system. Various types of communication of interest to NASIC are noted in this section and there is discussion of the use of direct dial-up versus dedicated communication lines.

A serious problem will tend to develop as a NASIC operation grows, in terms of the volume of items handled - the interconnection problem at its headquarters processing facility. It is proposed that for this potential interconnection problem, the possibilities of cable technology using time-division multiple-access digital communication in a high band-width "bus" channel be kept in view.

L. Technology Change and Information Service Center Planning (pp. 141-150)

Because NASIC is beginning operation during a period when computer and communication technology is undergoing rapid change, this section is in pp. 141-147 devoted to a brief overview of relevant technology with a summary of what this represents for NASIC operations and what the implications are for NASIC planning. Some of the topics discussed in this section are: trends in computer technology, integrated network systems, communications, large memories, low-cost processors, low-cost terminals.

It is emphasized that NASIC should set up certain experimental subsystems where possible, to provide a reasonable interplay with evolving technology. Consideration should be given to the use of NASIC as a test-bed situation where new developments could be tried and evaluated. It is doubtful that a NASIC operation based on "state-of-the-art" methods and means could provide coverage, speed of response and quality of service especially if a sizable user community is to be served.

M. Some Management Aspects of Information Service Center Operation (pp. 151-153)

In this section of the report there is discussion of several management aspects of interest to NASIC: objectives for operational support, objectives for the user market, kinds of service and promotion. An operation that will attract users who are capable and willing to pay for services, a market

SUMMARY Cont'd.

that includes users from the industrial and commercial community, a service that attempts to understand what the information need is and then focuses a means to meet that need, a promotional program that vigorously publicizes NASIC services and resources, are all desirable objectives which NASIC management will wish to consider and attempt to achieve.

A. NASIC OBJECTIVES AND SOME FACTORS INFLUENCING THE ATTAINMENT OF THESE OBJECTIVES

In broad terms, the function of NASIC is to improve the accessibility of information needed by the science community in the Northeast region of the United States, a major segment of this community being composed of science professionals associated with universities and colleges in the ten state area.* This goal could be stated more precisely as a pair of closely related objectives.

1. To improve the accessibility of information relevant to the Northeast science community.
2. To improve the exposure of the Northeast science community to relevant science information.

These objectives really form two sides of the same coin. The former implies the ability to make specific, relevant information and data readily available when the need for it arises. That is, in response to a specific request made, NASIC will initiate the searching of one or more appropriate data bases in order to deliver to the requester the data, documents or document references that appear to satisfy his immediate need. This type of service is frequently referred to as a "demand" or "on demand" service; it involves the searching of accumulated files of data or references (retrospective searching). The second broad goal, improving exposure, implies a service more dynamic in nature. In this type of service the information

*Maine, New Hampshire, Vermont, Massachusetts, Connecticut, Rhode Island, New York, New Jersey, Pennsylvania, Delaware.

center does not wait for people to approach it with specific demands. Rather, it attempts to inform users on a continuing basis of newly published literature of direct relevance to their current professional interests. This type of current awareness or alerting service is perhaps best exemplified by programs for the selective dissemination of information (SDI), in which a computer is used to match the "interest profiles" of users against the characteristics of documents recently added to a particular data base.

In both of these broad activities, NASIC will play a similar role. The role is essentially (see Figure 1) that of an interface between the scientific community and the universe of available bibliographic and data resources. The major component of these resources, and the component upon which we will concentrate in this report, will consist of data bases available in machine-readable form. However, in its overall service function, NASIC will certainly need to draw upon other resources; i.e., manual files of one type or another.

It is clear from Figure 1 that NASIC is involved in two interfacing operations. On the one hand, it interfaces with the members of the user population; on the other hand, it interfaces with the producers and suppliers of data bases (information wholesalers) and/or with various middlemen (information retailers) who already offer services from these machine-readable files.

In terms of effectiveness, the goal of NASIC is to provide the maximum level of exposure/accessibility to relevant

NASIC INTERFACE FUNCTION

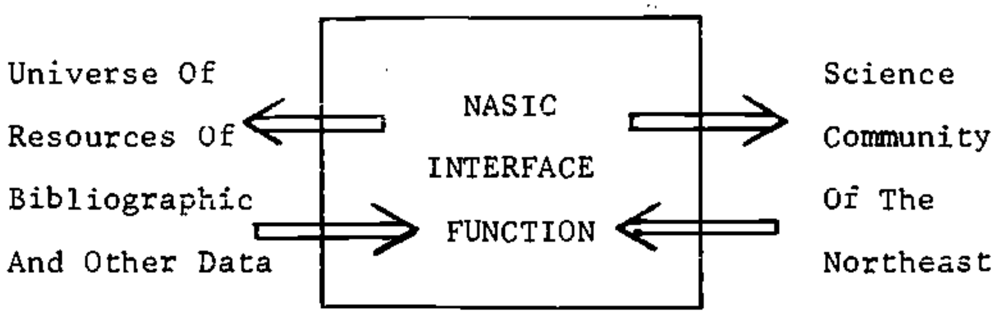

FIGURE 1

data bases for the science community in the Northeast. In terms of cost-effectiveness, the NASIC goal is to achieve maximum exposure/accessibility per dollar expended. A higher level objective is to provide to the science community the maximum possible benefits from the exploitation of information resources within the science community's available operating budget (i.e., cost-benefit considerations). Unfortunately, as pointed out elsewhere by Lancaster[1], the benefits of information are notoriously difficult to measure. Moreover, in actual practice, the distinction between cost-effectiveness analysis and cost-benefit analysis, while real, is not always absolutely clear and is therefore difficult to make.

This report will be concerned largely with various considerations relating to the effectiveness, cost-effectiveness and (ultimately) the benefits of NASIC operations.

Returning to the simple conceptualization of Figure 1, it is possible to identify three broad sets of factors that will exert a major influence on the performance of NASIC. These factors are:

1. Which data bases NASIC makes available, or more particularly, NASIC priorities for the acquisition of data bases and for the level of accessibility provided to the data bases.
2. How these data bases are "acquired" by NASIC. Here we are concerned both with business considerations (most convenient and inexpensive modes of acquisi-

tion of data base or data base services) and with technical performance considerations (e.g., a particular file may be searchable by a number of different searching systems and one may be more efficient than the others).

3. How the data bases acquired by NASIC are made available to the potential user population. Here our concern is with the products and services of NASIC, and how these services are presented to the user community. Pertinent considerations include possible modes of access to NASIC services, including interactions between NASIC representatives and the user communities, as well as telecommunications and networking considerations.

Since NASIC is directly concerned with the provision of information services, it is appropriate to give some consideration here to the factors that importantly affect the performance of an information service operation. These factors, which must be borne in mind in later decision processes relating to selection of data bases and modes of access provided to these data bases, are discussed in the next section.

B. FACTORS AFFECTING THE PERFORMANCE OF INFORMATION RETRIEVAL AND DISSEMINATION SYSTEMS

From the viewpoint of factors affecting system performance it is convenient to identify two types of information retrieval operation:

1. The delegated search. This is the situation in which the person needing information (i.e., the scientist or practitioner in some other professional field) delegates the responsibility for finding this information to a second person, usually a librarian or other information specialist. This mode of searching is the only one possible in machine-based retrieval or dissemination systems operated in an off-line, batch-processing mode.
2. The non-delegated search. This is the situation in which the person who needs information conducts his own search directly This mode is exemplified by a scientist's use of a printed index (e.g., Index Medicus) or by his interrogation of a remote data base by means of an on-line terminal.

The major factors influencing the success of a delegated search, with special reference to a search in a mechanized retrieval system, are depicted in Figure 2. Whether or not a requester approaches a particular information system or center in the first place is dependent upon his expectations regarding the scope and coverage of the service. Presumably he will not

FACTORS INFLUENCING THE DELEGATED SEARCH

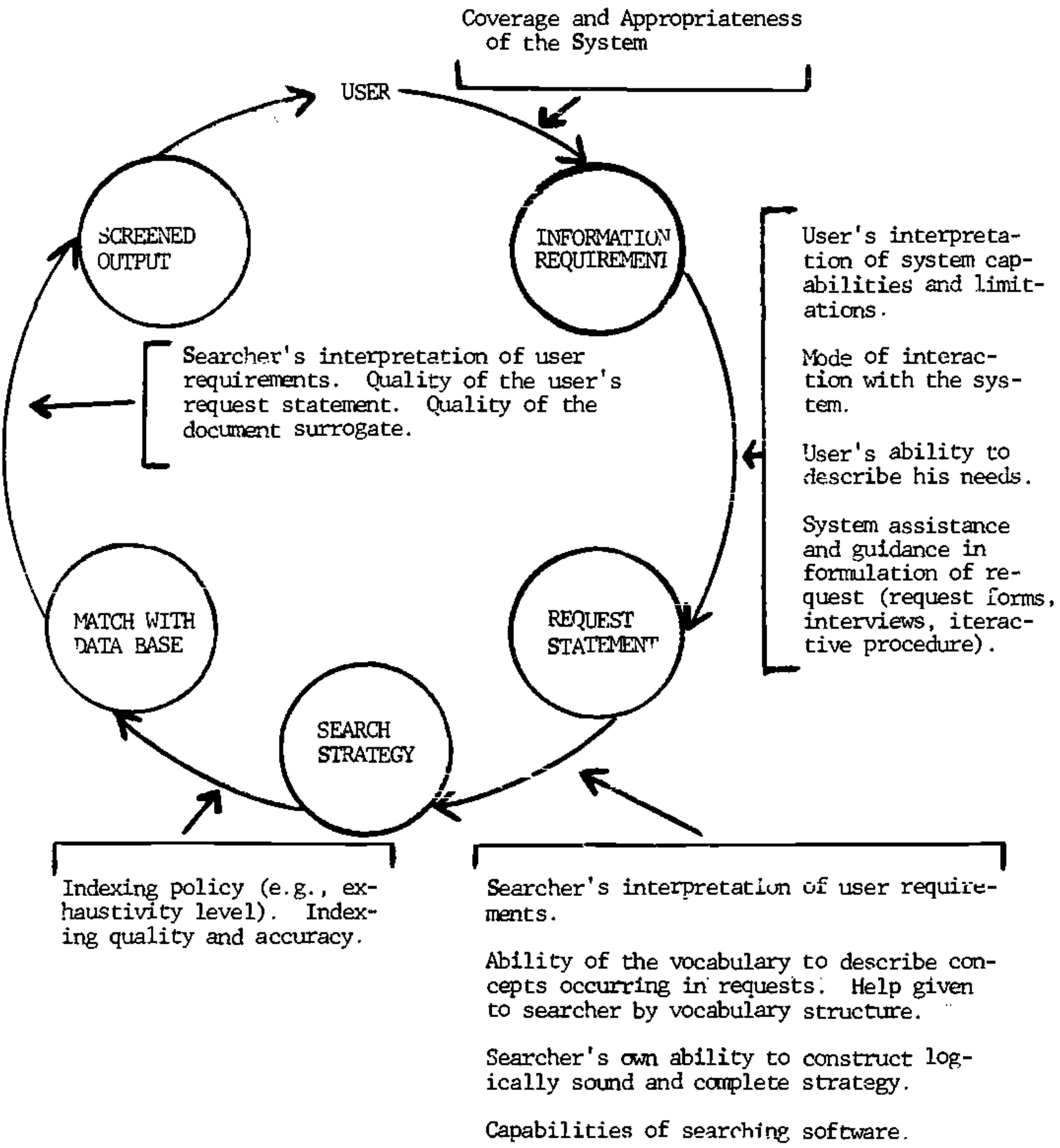

FIGURE 2

approach the system unless he feels that the collection is likely to contain the type of information or data he is seeking. Having decided to consult the system, he must make his needs known by means of a verbal request. The quality of this request (i.e., the degree to which it actually matches his information requirement) is dependent upon:

1. His interpretation of system capabilities and limitations. There is a strong tendency for a user to ask for what he thinks the system can give him rather than to ask for what he is really looking for.
2. His mode of interaction with the system.
3. His own ability to describe his needs, to express himself.
4. The degree of assistance and guidance given to the requester by the system. Such assistance can take various shapes: a carefully structured search request form, a formal "request interview" process, an iterative search procedure, or some type of user training program.

The request having been made to the system, it must be translated into a formal search strategy by a member of the information staff (search analyst). Now a new series of variables, affecting the recall and precision of the search, come into play:

1. The analyst's own interpretation of what the user really wants (which may be accurate or inaccurate).
2. The ability of the vocabulary to express the user's need. For example, the user may specifically be seeking articles on "argon arc welding"(and the search analyst recognizes this) but the vocabulary may only be capable of expressing this at a higher generic level - "shielded

arc welding" or "arc welding" - and thus precision failures are inevitable.

3. The ability of the search analyst to recognize and cover all possible approaches to retrieval. To take a simple example, the requester may be looking for articles on possible adverse effects of commonly consumed beverages or components thereof. The searcher uses the terms "caffeine", "coffee", "tea", and "theophylline", but forgets about the possibility of "cacao" and "theobromine" and thus misses some of the relevant documents.

   The "level"of search strategy adopted. The searcher can choose to use a broad strategy (leading to high recall but low precision) or a tight strategy designed for high precision (but usually at the expense of a low recall) or a compromise between the two extremes.

4. The capabilities of the searching software.

When the search strategy is actually matched against the data base (i.e., the search is conducted), further factors affecting performance come into play. One important performance factor is that of indexing policy, particularly policy regarding exhaustivity of indexing (which really equates with the number of index terms or other access points provided). Perhaps the exhaustivity of indexing is inadequate to allow some of the relevant items for a particular request to be retrieved. Inaccuracy of indexing (omission of important terms or assignments of terms incorrectly) will also lead to recall or precision failures.* The characteristics of the vocabulary affect the indexing process as much as they affect the searching process. An indexer can only adequately represent the concepts occurring in a document if there are appropriate specific terms available for him to use. Lack of specificity in the vocabulary will usually cause precision failures but can also lead to recall failures. Further, the vocabulary must be capable, to a certain extent, of showing the syntax of the terms assigned in indexing, thereby avoiding at least some of the precision failures that would be caused by false coordinations or incorrect term relationships.

Finally, before the results of a search are submitted to the requester, the analyst may screen the output and eliminate items that appear to be irrelevant with the object of improving the precision of the search to the end user. How successful this screening operation is (i.e., how much precision can be improved without having too serious an effect on recall) depends primarily upon the accuracy of the analyst's interpretation of the requester's requirements. Secondarily, the success of the screening will be affected by the quality of the document surrogate from which the analyst is working.

Of course, these various sources of failure are cumulative. For a particular search conducted in a retrieval system, some of the relevant documents may be missed by the very fact that the user's request statement is too restrictive and inadvertently ex-

---

*A recall failure is the failure of the system to retrieve a relevant document or other item. A precision failure is the reverse of this, the failure of the system to avoid an irrelevant item.

cludes certain items. Others may be missed due to poor search strategy, vocabulary inadequacies, indexing policy, and indexer omissions. Finally, the analyst may eliminate some more relevant items in his screening process. With so many possible sources of loss, it is little wonder that systems do not on the average operate very close to 100% recall. A similar cumulative effect occurs to prevent us obtaining 100% precision.

The performance factors illustrated in Figure 2 are relevant to all types of delegated searching systems, manual as well as mechanized, dissemination systems as well as retrospective searching systems. It is important to keep these factors in mind when making decisions relating to the selection of data bases and the provision of services from these data bases.

It is clear, for example, how the characteristics of the data base itself exert a considerable influence on the performance level of services provided from this data base. In the first place, in its subject coverage and level of treatment, the data base searched must be appropriate to the information need for which the search is being conducted. Selection of the most appropriate data base to search is clearly a first step that is critical to the entire information seeking process. In interrogating a particular data base, the searcher is limited in what he is able to do by certain inherent characteristics of the data base itself. The most important of these characteristics are the <u>exhaustivity</u> of the indexing and the <u>specificity</u> of the vocabulary used in indexing. An <u>exhaustive</u> representation of a document is one that presents a fairly complete searchable representation of its subject matter. The most exhaustive "representation" of a document is the complete text of the document itself, available in machine-readable form for searching on a word-by-word basis (as, for example, in certain legal retrieval systems). A fairly full representation would consist of a searchable abstract or a set of, say, 10 - 20 index terms which collectively represent at least the major subject matter of the document. A record of low exhaustivity would consist only of the title of a document or a small number of index terms representing only part of the subject matter covered. By "exhaustivity of indexing", then, we mean essentially the number of access points provided to a bibliographic record. Exhaustivity is the prime factor governing the <u>recall</u> capability of a retrieval or dissemination system (i.e., the ability of the system to retrieve relevant references) since it is clear that any particular record can only be retrieved by the access points provided in the data base. That is, a data base in which the records are exhaustive representations of the subject matter of documents is capable of providing a high recall whereas a data base in which the records are nonexhaustive representations of subject matter cannot provide a high recall except at the expense of an unacceptable level of precision (to retrieve a high percentage of all relevant items on a particular topic we would need to retrieve a very large number of items most of which would not be relevant to the specific subject of the search).

The precision of a retrieval system is largely governed by the specificity of the vocabulary used to index subject matter. A highly specific vocabulary will allow a high level of precision in searching whereas a nonspecific vocabulary dooms a system to low precision. That is, a searcher cannot interrogate a data base any more specifically than the vocabulary of the data base allows him to. For example, a searcher in MEDLARS may be looking for information on acute frontal sinusitis. The vocabulary of the system does not permit indexing of this subject matter at this level of specificity. The most specific relevant index term is "sinusitis", which means that the entire class of documents on "sinusitis" must be retrieved in a search on "acute frontal sinusitis". Since most "sinusitis" documents will not be relevant to the specific topic of "acute frontal sinusitis", it is obvious that a search on this topic will have a low precision. In general, natural language data bases (e.g., those providing a complete searchable abstract) provide high specificity and therefore are capable of high precision in searching. Controlled vocabularies, such as thesauri or lists of subject headings, tend to be less specific than natural language, and thus tend to operate at a somewhat lower level of precision. On the other hand, natural language data bases may create greater problems of semantic and syntactic ambiguity. These problems take two forms: false coordinations (the situation in which two terms on which a search is conducted, while present in a retrieved document, are essentially unrelated in that document) and incorrect term relationships (the situation in which two search terms, although related in a document, are related in a way other than the relationship sought by the requester - e.g., the difference between "reading" stimulating "epilepsy" and the "reading disabilities of epileptic children").

Two further data base characteristics exert influence on the performance of a retrieval or dissemination operation. The first of these is the quality of the indexing. Whereas exhaustivity of indexing is controlled by policy decisions made by system managers. the quality of the indexing (or lack of it) is controlled by the individual members of an indexing staff (assuming that some human indexing operation does actually take place). Indexing errors are of two types:

a. The omission of an important index term, or

b. The use of an incorrect term, i.e., a term inappropriate to the subject matter covered by a document.

If a data base is well prepared, with some built-in quality control procedures, indexing errors are likely to be relatively infrequent. But, some machine-readable data bases are far from error-free and their effectiveness is diluted as a result.

The second characteristic relates to the "indicativity" of the record stored in the data base. It is important that the document record provided by a retrieval or dissemination system (in a printout or in an on-line display) gives a representation of its subject matter sufficient to allow a reader to determine its probable relevance to his own information need or that of someone else for which the search is being conducted. In general, titles are inadequate indicators of subject content; abstracts or the complete set of index terms assigned to a document (i.e., the "tracings") are somewhat better. Studies conducted by Project Intrex (see Marcus et al (2)) discovered that relevance decisions made on the basis of document titles agreed at the 60-70% level with relevance decisions made on the basis of the complete text of these documents, whereas relevance decisions made on the basis of abstracts or lists of index terms agreed with those made on full text at the 70-90% level. Similar findings have been made by Saracevic (3). The indicativity of a record is roughly proportional to the length of the record in English words. That is, the longer the record the more useful to the user it is likely to be as a predictor of relevance, but, of course, the more costly the record to store and to print out.

Turning once more to Figure 2, it can be seen that there are two further broad categories of factors influencing the performance of an information retrieval system. These factors, not directly related to the data base itself, are related to the searching strategies used and the method of interaction between the user of the system and the system itself.

In any delegated search system a major factor influencing performance is the mode of interaction between the user and the system in the negotiation of the user's actual information need. The result of inadequate interaction is that the search analyst (or other information specialist) is left with an imperfect representation of what the user really wants. Under these conditions a fully successful search is unlikely, and it is quite possible that the results will be completely unacceptable to the register. The major factors influencing the success (or failure) of the user-system interaction process are illustrated in Figure 2 and were touched upon earlier. The problems of user-system interaction in the negotiation of search requests have been discussed in detail by Lancaster (4)(5). In his evaluation of MEDLARS, Lancaster discovered that imperfect user-system interaction accounted for about 25% of all the recall failures and 16% of all the precision failures occurring in the system. Thus, this was a major source of system failure as it will always tend to be in a machine-based delegated search system. Clearly, this has important implications for the procedures whereby the NASIC information services librarians interact with the user community, a point we will return to in greater detail later.

The final set of factors influencing the performance of a retrieval or dissemination system are the searching strategies themselves. As indicated in Figure 2, the quality of a search strategy is dependent upon a large number of factors, including the searcher's interpretation of user needs and his own ability to construct a strategy that is complete (i.e., covers all possible approaches to the subject matter) and logically correct. However, the searcher is somewhat constrained by various data base characteristics and by the properties of the searching software available to him. Two important data base constraints have been mentioned already: the exhaustivity of the indexing (which essentially governs recall) and the specificity of the vocabulary used in indexing (which, in an absolute sense, controls the level of precision possible). It is not possible for a searcher to overcome these data base constraints; that is, the degree to which he can vary a search strategy to produce a high level of recall or a high level of precision is largely dependent upon the exhaustivity of the indexing and the specificity of the vocabulary. One major task faced by a searcher in a machine-based system operated off-line is that of thinking of all possible approaches to the retrieval of material on the subject of the search. How successful he is in this is at least partly dependent upon data base characteristics again; that is, he is more likely to be able to devise a comprehensive strategy if the data base employs a highly structured vocabulary that explicitly reveals hierarchical and other semantic relationships among terms. The thesaurus of such a system provides a very important aid to the searcher.

Another constraint is placed upon the searcher by the searching software available. Obviously, the searcher is limited by the capabilities provided in this software (e.g., truncation capabilities, word proximity operators, nesting levels, and so on). The more flexible the software, the more features it has, the more powerful is the tool available to the intelligent searcher. Clearly these various factors must be borne in mind by NASIC in the selection of data bases, the selection of service centers and/or the selection of available software for in-house use.

While Figure 2 depicts the major factors influencing performance of a delegated search system (as exemplified by one operating in an off-line, batch processing mode), Figure 3 depicts the major factors affecting the performance of a nondelegated search system. Figure 3 would, for example, be typical for an on-line retrieval system interrogated directly by a scientist to satisfy his own information needs. This is in some ways a more simple situation. The "interaction" failures that tend to be prevalent in delegated search systems are avoided, but the vocabulary and indexing factors are just as important in the on-line situation as they are in the off-line. Apart from these data base constraints the major factors influencing the success or failure of a nondelegated on-line search are the searcher's knowledge of data base characteristics (especially his knowledge of the vocabulary of the system), his ability to verbalize his requirement and to think of all reasonable approaches to retrieval (assuming that he wishes to achieve a high recall), the assistance and guidance given by the system, his ability to construct logically correct searching strategies, and the capabilities of the searching software he is working with.

If the on-line system is being used in the delegated search mode the situation is much the same as the situation in the off-line, batch processing search as presented in Figure 2. There is one important difference, however; the off-line search is noninteractive and nonheuristic. The searcher must think in advance of all likely approaches to retrieval. Mistakes are relatively costly since the searcher will pro-

FACTORS AFFECTING THE NON-DELEGATED SEARCH

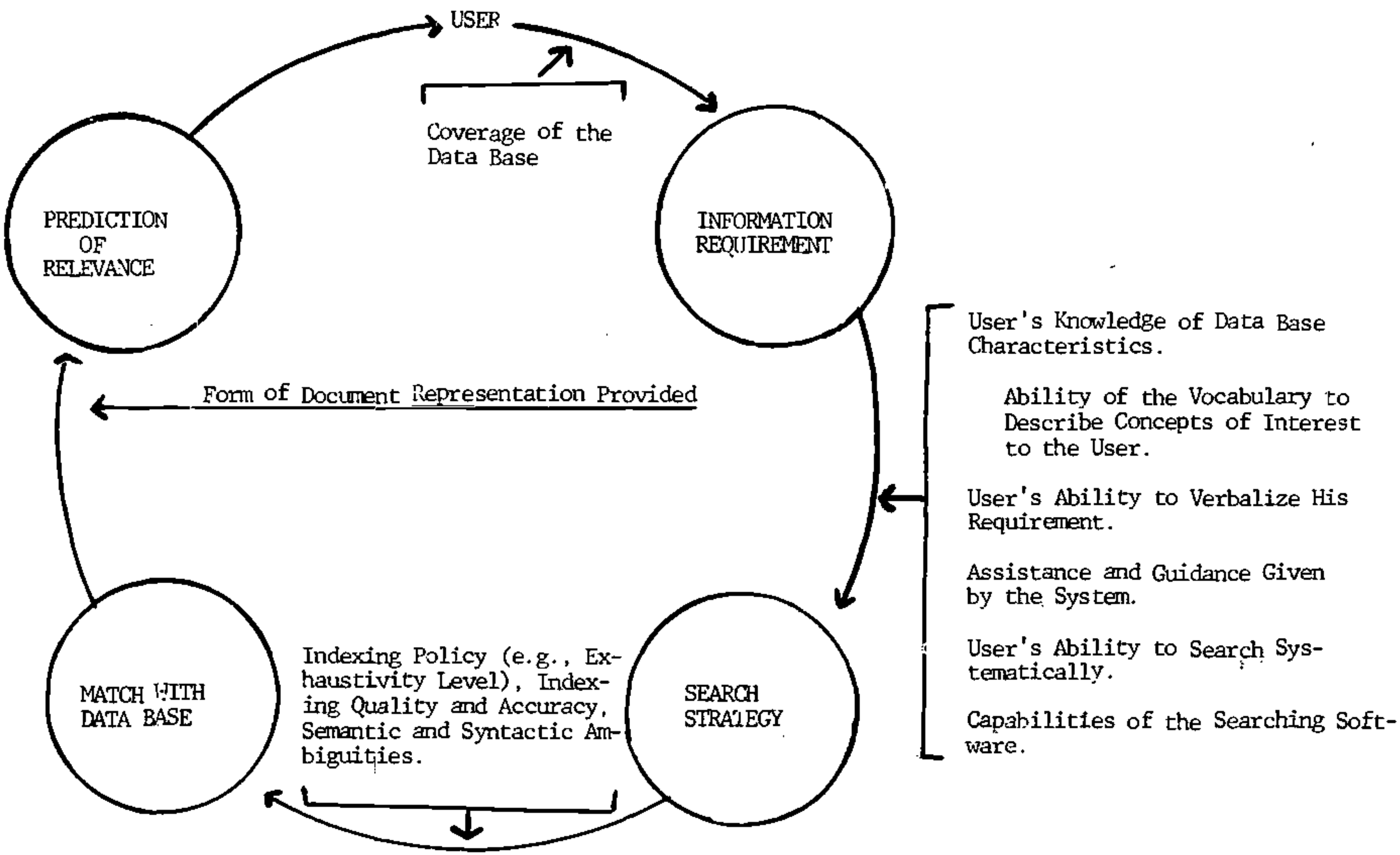

FIGURE 3

bably not know for some days(i.e., until he receives his search printout) whether or not his search has been successful. If it has been unsuccessful, he must start all over again. In the on-line search, on the other hand, he can see immediately whether or not a particular strategy is "hitting" the type of item he is seeking. If not, he can modify his strategy immediately. Searching on-line is interactive and heuristic and mistakes are less costly because they are easily noticed and corrected. A certain amount of "browsing" is also possible in an on-line system.

In all of this discussion we have concentrated on bibliographic systems at the expense of data retrieval systems. This is deliberate. Bibliographic systems are much more complicated in that they are dealing with "soft" rather than "hard" data. In searching for documents (or references to them) the searcher is faced with problems of semantic ambiguity and the inherent imprecision of language. There is no answer from a bibliographic system that is inherently and unequivocally "correct". On the other hand, a system dealing with numerical data would generally be free from semantic ambiguity. There is usually one correct answer to a query posed to such a system (i.e., some numerical value).

Nevertheless, some of the problems identified in relation to the use of bibliographic systems may be equally relevant to the use of data retrieval systems. The user of a data retrieval system will also be constrained by data base characteristics, including the amount of data provided and how this data is formated. The user will also be concerned with how comprehensive the data base is, how accurate the data is, and how current it is. Another important consideration will relate again to searching software; that is, just how can the data be manipulated by the software available?

We have gone to some lengths to discuss these factors that importantly affect the performance of all information retrieval and dissemination systems because it is important that all of these things be kept in mind in making decisions relating to NASIC operations. Clearly, we should try to identify the major factors likely to critically affect the performance of NASIC information services. Knowing what these factors are will help in making decisions relating to data base evaluation and selection, data base acquisition, modes of interaction with NASIC users, and products and services to be provided. Some guidelines relating to these various decision processes will be discussed in the ensuing sections of this report.

C. CRITERIA RELATING TO THE SELECTION OF DATA BASES

A major emphasis in the NASIC program will be the provision of services based upon machine-readable data bases. Less than a decade ago, almost nothing of significance in the way of bibliographic or numerical data bases was available in machine-readable form. The situation has changed dramatically. The first general survey of available files, that of Carroll (6), listed 55 machine-readable files, and an Auerbach report (7), listed 34 suppliers of machinable bibliographic records, both of the lists appearing in 1970. An ASLIB survey of 1972 (8) identified 48 different services, with special reference to services available in the United Kingdom. The ASIS survey (9), published in 1973, provides information on 81 commercially available machine-readable files, while a list issued by the LARC Association (10), also in 1973, mentions 122 "available data banks for library and information services". Probably the most complete survey, however, was issued by Computer Sciences Corporation (11) in 1972. This survey mentions 169 information systems having machine-readable data bases, although 268 such systems had actually been identified by the compiler. Conservatively, then, one could state that there exist on a worldwide basis at least 300 machine-readable bibliographic files, although not all of these are readily available for outside use. If we add data files to this, the total number of data bases will be very much greater.

With so many data bases to choose from, how can NASIC, or any other information center, decide which to make available and in what order of priority? Disregarding any contractual constraints, it is useful to give some general consideration to the subject of evaluation of available data bases. Some of the most important selection criteria are summarized in Table 1.

Without any doubt, the major consideration is the subject matter of each data base. The major concern of NASIC, or any other information center, should be to make services available in order of priority according to the anticipated demand for these services within the user community to be served. This implies the need to match data base content with the "subject profile" of the NASIC user community. In terms of disciplines and major subdisciplines such a profile can be constructed from faculty directories, college catalogs, membership lists, and similar publications. This type of broad analysis will yield the approximate numbers of potential users of information services in general subject areas: chemistry, physics, medicine, electrical engineering, education, and so on. To assess the potential market for interdisciplinary services (e.g., the NASA data base or a data base in the area of environmental protection), for services of a more specialized nature (e.g., in spectroscopy, in toxicology, in crystallography), or for services covering particular types of material or data (e.g., patents, product catalogs) is rather more difficult and will involve the use of more sophisticated market analysis techniques.

TABLE 1

CRITERIA AFFECTING THE SELECTION OF MACHINE-READABLE DATA BASES

1. Subject Matter Of Data Base

Match of subject matter with subject interests of user community.

2. Cost Factors

Cost of "acquiring" data base or services from it.
Unit cost per user interest profile.
Unit cost per retrospective search.
Unit cost for group profiles.
Cost of acquiring data base in relation to number of records provided.
Cost of acquiring data base in relation to number of access points provided per record.
Cost of data base in relation to "quality" considerations.

3. Quality Considerations

Coverage

Coverage by number of sources.
Coverage by type of source.
Coverage by number of "items".
Coverage by time span.
Completeness in relation to specific topics of interest to user community.
"Uniqueness" of data base. Overlap with other data bases.

Time Factors

Time lag in inclusion of sources.
Frequency of file update.

Indexing and Vocabulary Factors

Specificity of vocabulary.
Is vocabulary controlled?
Are searching aids provided?
Prevalence of semantic and syntactic ambiguity.
Exhaustivity of indexing. Number and variety of access points provided.
Accuracy and consistency of indexing. Observed error rates.

4. Implementation Factors

Assurance of continuity.
"Cleanliness" of data base.
Compatibility with in-house software and hardware.
Amount of pre-processing needed.
"Integratibility" with other data bases handled.
Capabilities of searching software available.

The cost of making a particular data base available must be considered in relation to anticipated demand for that data base. Some of these cost considerations will be considered in greater detail in the next section of this report. In general, the unit cost of any information service is extremely volume-dependent. This would be particularly true in the case of a data base that must be purchased or licensed directly and manipulated in-house on NASIC facilities. It will also be true, although to a lesser extent, in the purchase of services through an existing information retailer. That is, NASIC can and should expect to be able to negotiate reduced rates for a high volume of use of any particular service. Since extent of use of a particular service will be at least partly dependent upon the cost to the user, NASIC must endeavor to estimate a realistic level of potential demand for each service and calculate the unit cost of this service (e.g., per retrospective search, per SDI profile) based upon this expected level of demand. Top implementation priority should be given to those data bases that, because of anticipated demand and estimated unit cost per service unit, are likely to attract the greatest volume of business to the information center.

Independent of expected demand, however, there are other cost effectiveness considerations that might be taken into account in the evaluation of machine-readable data bases. If a data base must be purchased, leased or licensed, the purchaser should consider what exactly he is getting for the money invested. One possible measure of return on investment is the cost per searchable record. For example, the Science Citation Index tapes can be leased (source tape plus citation tape) for $20,000 per year. This buys access to approximately 400,000 searchable records, an average cost of about 5 cents per year per searchable bibliographic record. In contrast, it costs $30,000 per annum to lease one year's worth of MEDLARS tapes, containing about 250,000 records, which gives a cost of rather more than 8 cents per record.*

The size of the data base is not, of course, the only consideration. Another important cost-effectiveness factor would be the number of access points provided per record since this, ultimately, controls the retrieval capabilities of the system. A data base providing a large number of access points per record (e.g., a citation plus complete searchable abstract or a citation plus an average of 10 - 20 humanly assigned descriptors) will generally be more costly to create and thus, presumably, more costly to acquire by purchase, leasing or licensing arrangements. Such a file would also be more difficult to compress and thus more costly to store and manipulate.

---

*It seems that most data bases lease for something in the range of 5 to 10 cents per item per year and that the leasing rate should generally not exceed this. Leasing rates for the Geo-Ref files of the American Geological Institute work out to about 10 cents per record per year. For the METADEX files of the American Society for Metals the cost is 5 cents per record per year (i.e., for $1250 about 25,000 items are available through the leasing arrangement).

Nevertheless, in such a file, each record may be accessible from a large number of different approaches. Since, essentially, an information center acquires a data base in order to obtain convenient access to the individual items in this file, the cost of offering service from a particular data base needs to be balanced against the amount of access the file provides. From a cost-effectiveness viewpoint, the price of a data base must be related not only to the number of records the investment is purchasing but also the amount of access to the records that is being purchased. As an example, the MEDLARS data base, or a large portion of it, is indexed quite exhaustively. The file is expensive to lease but this cost is not necessarily excessive in relation to the amount of access provided.

So far we have talked only of quantitative considerations - how many records are purchased, how much access is purchased. The quality of the data base must also be considered, although quality is not always easy to assess. Major qualitative considerations would include the following: (1) Coverage, (2) Time Factors, (3) Indexing And Vocabulary Considerations, (4) Continuity, and (5) Ease Of Implementation.

Coverage

Coverage is a critical consideration and there are a number of dimensions to this problem. One we have already mentioned is the number of individual items included in the file. A second is the scope of the coverage. Scope may be expressed in terms of the number of different services regularly covered in the data base (e.g., number of journals indexed) and in terms of the range of types of sources included (journal articles, technical reports, patents and so on.) The National Library of Medicine, for example, regularly indexes close to 250,000 items a year from about 2500 different sources, but this coverage is almost entirely from biomedical and general scientific journals. The coverage of technical reports is very limited, as is the coverage of conference proceedings or symposia (except those appearing in a regular series) and medically related patents are not included at all.

Another dimension of data base coverage relates to time span. To be of any use at all, for most purposes of retrospective search, a data base must span at least 2 - 3 years of the literature, and it is probably only beginning to approach a very high level of value when it covers about five years of literature. Hansen (21), for example, reports little interest from users in a search of CA Condensates tapes going back 2 - 3 years only. In the humanities and social sciences time span is somewhat more important than it is in scientific and technological fields, where literature and some data tend to be superseded rather rapidly.

It should always be borne in mind that the great discipline-oriented services are not necessarily complete, or even close to complete, in specific topical areas falling legitimately within their own scientific disciplines. Davison and Matthews (12), for example, found that twelve major indexes relating to chemistry and spectroscopy each covered only a very small proportion of a collection of 183 references known to exist on the topic of "computers related to mass spectrometry". In fact, no single source included more than 40% of the known references. Chemical Abstracts was found to include only 24% of these references, Referantivyni Zhurnal Chimii only 5.5% and Chemisches Zentralblatt only 4.5%. For this reason it is important that NASIC services include the widest possible use of specialized information and data centers, whether they have machine-readable files or not, as well as the great discipline - or mission-oriented data bases. In the area of mass spectrometry, for example, it would be important to use the services of the Mass Spectrometry Data Center (of the United Kingdom Atomic Energy Authority) as well as Davison's own service at the Scientific Documentation Centre, Dunfermline, Scotland.

Bourne's (13) evaluation of the Bibliography of Agriculture concluded that this source included only 48 - 58 percent of the available literature relevant to the interests of agriculture researchers and that the material missed was not at all of an evanescent nature but was predominantly English language material, mostly from journals and conference proceedings, and much from U.S. sources. An analysis of the coverage of material from USDA, a research laboratory, and several state agriculture experiment stations or extension services indicated that the Bibliography of Agriculture appeared to cover only 45 to 74 percent of this type of material. An extensive program of studies by Martyn (14) also indicated that it is unrealistic to expect that any one service relatively broad in scope is likely to be comprehensive in its coverage of some specific subtopic falling within the broad subject area.

Indeed, it would be very unusual if a broad discipline-oriented service did provide comprehensive coverage of some specific subtopic. The "law of scattering", first propounded by Bradford (15) in 1935 and since substantiated by numerous other investigators (see Fairthorne (16)), indicates clearly, that, while a relatively small number of "core" services are likely to account for a relatively large proportion of all the references on a particular topic, the remaining references will be very widely dispersed over a great many sources, including sources that are very peripheral to the subject area covered by the core journals. In other words, the distribution of references over sources is hyperbolic in nature. This "empirical hyperbolic distribution" is identical with the distribution of the use of words in published text, as observed by Zipf (17).

It is likely that a number of different data bases may need to be used to achieve a really comprehensive coverage of literature on some particular topic. For example, Montgomery (24) found that four data bases collectively covered 98% of a sample of toxicology articles published in 1968, but the maximum coverage of the first service was 85%. Another factor relating to coverage is the "uniqueness" of the data base. There is a considerable amount of overlap among several major data bases. For example, there is great duplication between MEDLARS and EXCERPTA MEDICA and it has been discovered that at least 50,000 articles a year are abstracted by at least two of the services BIOSIS, CAS and Engineering Index (18), although this duplication is likely to be reduced in the future by means of cooperative arrangements between the publishers of these services. Certain other services (19) on the other hand, are relatively "unique"; that is, they overlap others very little. Bourne (20), for example, discovered that no other service overlapped the Bibliography of Agriculture by more than 20%. Duplication is not in itself necessarily bad. Because of the interdisciplinary nature of many subject areas one would expect, in striving for completeness, that one major service might overlap another. Nevertheless, in establishing priorities for the implementation of services, NASIC should give careful consideration to matters of overlap and uniqueness and should endeavor to implement as early as possible a set of data bases that collectively give the broadest possible coverage of the entire scientific and related literatures.

Before we leave this matter of coverage two other factors should be mentioned. First, it is important that a service covers regularly and consistently all the sources it claims to cover. In other words, the reliability of the coverage of sources is important. We must be able to rely on a service to cover all issues of a particular journal, all reports of a particular series, and so on. If we cannot rely on a service in this way (i.e., we find that it sometimes indexes a particular journal and sometimes does not) its value is seriously degraded.

Another factor relates to the reliability of the service in selecting all articles, from the sources it claims to cover, that legitimately fall within the subject area of its coverage. A comprehensive service such as the Science Citation Index will include everything from the sources it covers, but other services, while they may cover some sources comprehensively, will cover others selectively. Again, if we cannot rely on the service to pick up everything of relevance to its stated scope, its value will be degraded. Devon et al (23), for example, in comparing Ringdoc with CBAC, found that the latter covered 580 journals to the former's 332 but produced evidence that suggests that Ringdoc's selection of articles is better, at least in the pharmaceutical area.

Stern (25), after disclosing that no single source is likely to give a comprehensive coverage of all literature relevant to the pharmaceutical industry, suggests that it should be possible to plan an information retrieval strategy in terms of the cost needed to obtain a particular level of coverage. It is possible to select one or more services which, in a particular subject area, will yield a given level of coverage (70%, 80%, 90%). An alternative strategy would be to select those services that are likely to give the best coverage for a specified amount of money available.

Time Factors

Another quality factor relating to a machine-readable data base is the extent to which it is up-to-date. The time lag between publication of a paper or report and its appearance in an abstracting and indexing service will have a significant effect on the value of that service, at least for current awareness purposes. In general, services in which human intellectual operations have been eliminated or at least kept to a minimum (as in the case of the Science Citation Index) are likely to be more current, all other things being equal, than a service in which delays occur in human indexing or abstracting operations. Even two services based on human indexing may exhibit substantial difference in their degree of currency. Bourne's study of the Bibliography of Agriculture (20), for example, showed that on the average, agriculture-related literature, included in this publication, appeared later than it did in eight of eleven services with which it was compared. Based on a sample of 617 citations published by both services, it was found that Chemical Abstracts, which includes abstracts in addition to the citation data provided in the Bibliography of Agriculture, published 3.7 months earlier on the average.

To be judged valuable by a scientist, a current awareness service must not only inform him of items relevant to his interests but it must also inform him of many of these items before they are brought to his attention in any other way. In other words, a current awareness service must have a high novelty factor if it is to survive as a commercial venture. A service is unlikely to survive if, due to processing delays, it brings to the attention of its customers literature that they were mostly aware of previously. In comparing the suitability of a particular data base for providing current awareness services NASIC must give this factor of currency careful attention. A data base that is not very current, at least for a major proportion of the sources it covers, should not be used for dissemination services, although it may still be quite valuable for retrospective search purposes. Given two data bases covering essentially the same subject area (e.g., MEDLARS and EXCERPTA MEDICA) the more current should always be given preference, all other things being approximately equal.

In evaluating various data bases in respect to their timely coverage of pharmaceutical information, Ashmole et al (25), discovered that the Science Citation Index data base (ASCA) averaged delays of only 0 - 3 weeks from time of publication, while most other services averaged 2 - 6 months, and Biological Abstracts averaged 4 - 12 months.

A related time factor to be taken into account is that of how frequently a machine-readable data base is updated. For current awareness purposes, a monthly update is probably the minimum that is acceptable, and an update every two weeks is preferable. For the purposes of retrospective search, frequency of update is somewhat less critical. Some suppliers of machine-readable data bases offer a differential pricing structure based on frequency of updating. That is, it is possible to lease or license a data base more cheaply if the purchaser will accept less frequent updating. Savings of this type are unlikely to be justified if the data base is one in heavy use or if it is used as the basis of dissemination activities.

Indexing and Vocabulary Considerations

These factors have already been discussed to some extent. The terms used to represent the subject matter of documents (whether there are natural language words in titles or abstracts, subject headings, descriptors, or codes from some classification scheme) must be specific. A nonspecific vocabulary will not allow the conduct of specific searches and a system based on broad terminology is doomed to a low level of precision.* Natural language vocabularies (i.e., uncontrolled) tend to be quite specific. For example, a searchable natural language abstract, if well constructed, is likely to be a highly specific representation of the subject matter of a document. Data bases in which subject matter is represented by index terms selected from a controlled vocabulary (e.g., a thesaurus) tend to be somewhat less specific. Thus, all other things being equal, a natural language data base is likely to provide the capability for achieving a higher precision than a data base using a controlled vocabulary. The capability of searching very specifically is more likely to be of importance for retrospective search services (where some requests

*It should be borne in mind, however, that the specificity of an indexing vocabulary must be considered in relation to all terms assigned to a particular item. A term standing alone may be quite general, but it may take on a specific connotation when used in conjunction with some other term.

may be made for quite precise topics of limited scope) than it is for current awareness purposes (where users are usually seeking coverage of broader areas related to a wider range of current professional interests).

Although natural language is generally quite specific, a data base in which only natural language (e.g., titles and/or abstracts) is used does present various other problems to the searcher. In a data base founded upon a controlled vocabulary, this vocabulary serves several important functions. It controls synonyms and near-synonyms, differentiates homographs, and links together terms that are semantically related. The thesaurus, or other form of controlled vocabulary, thus normalizes the language used in indexing and provides a very valuable aid to the searcher. Without some form of controlled vocabulary the searcher (or constructor of profiles) is "left to his own devices". It is up to him to think of all possible ways in which a particular topic might be represented in natural language titles or abstracts (not an easy task) because the system gives him no help. Actually, this last statement is not always true. Some systems based on natural language issue some type of searching aid in which the substantive words occurring in the data base are listed and grouped in various ways that may be of use to the searcher.

Semantic and syntactic ambiguities of the "false coordination" and "incorrect term relationship" type are likely to be more prevalent in the searching of natural language abstracts than they are in the searching of data bases in which subject matter is represented by index terms selected by indexers from a controlled vocabulary.

A second, related quality consideration is that (previously mentioned) of the exhaustivity of the indexing - the number of access points to subject matter provided. For comprehensive (high recall) searches an exhaustively "indexed" data base is a necessity. A data base in which a very limited number of access points is provided (e.g., titles only or only 2 - 3 index terms) is unlikely to be capable of providing a high level of recall at an acceptable level of precision. O'Donohue (22), describing experiences with a number of search services, points out that "references that were overlooked in computer scanning usually had inappropriate or insufficient keywords". In this connection it is interesting to note that searches of CA Condensates tapes at the National Technological Library, Denmark, have consistently been able to operate only in the range of 30 - 40% precision, even when special efforts were made to reduce irrelevancy (21) as much as possible.

A complete, searchable abstract is, of course, one form of exhaustive indexing. Multiple access points can be provided in a number of ways, including citation linkage. However, to ensure the high recall capability that may be needed in a current awareness or retrospective search system, the subject matter of a document needs to be represented directly and exhaustively in the document record stored in the data base.

Other quality factors relate to the accuracy and consistency of the indexing. Some machine-readable data bases have been found to contain a high level of indexing error and/or inconsistency. These problems can only be recognized through experience in using a particular data base or through ascertaining the experience of other data base users.

While NASIC will need to make its own evaluation of various data bases, in terms of the various criteria mentioned in this report and the recognized needs of the NASIC community, it can also draw upon the evaluations conducted by others. A few evaluations, some comparative, have appeared in print. Besides the study of O'Donohue (22), a comparative evaluation of Ringdoc and CBAC has been reported by Devon et al (23), and six different data bases were compared by Beauchamp et al (29) in terms of their yield in searching for literature on various chemical compounds. Scot et al (30) have reported on an evaluation of the Drugdoc service of EXCERPTA MEDICA, and Ashmole et al (25) have described a comparative study of several data bases in terms of their "cost-effectiveness" in searching for pharmaceutical references.

A final group of considerations relating to the acceptability of machinable data bases is concerned with reliability and ease of implementation. Most of the very large machine-readable bibliographic files are produced as a by-product of a publishing operation. This creates the danger that the records may contain various elements that are needed for publication purposes but are superfluous or possibly even obstructive in the use of the file for literature searching purposes. This problem of file "garbage" is less critical now than it was formerly since several publishers generate a separate "clean" file for search purposes. Nevertheless, some machine-readable records still contain a considerable amount that is "garbage" as far as the information service center is concerned.

Continuity

A major consideration for any service center should be the degree to which the continuity of the data base is assured. Of course, many suppliers of data bases have a long record of successful publishing behind them and it is not likely that these enterprises will "fold" in the forseeable future. In the evaluation of newer data bases, on the other hand, NASIC should pay special attention to the probable continuity of the file and should avoid the development of services relying on data bases whose future does not seem to be assured. Data bases compiled with the aid of funds supplied by government agencies, without other guaranteed sources of financial support, are especially suspect. As of this writing, for example, there is a strong move on the part of the National Institute of Neurological Diseases and Stroke to withdraw support from most elements of the Neurological Information Network, including the important services of the Brain Information Service (UCLA), which is estimated to actively serve some 20,000 neuroscientists on a worldwide basis.

Ease Of Implementation

Finally, in the case of data bases to be implemented by NASIC in-house, ease of such implementation must be considered. Considerations here will include record format, code structure (ASCII, EBCDIC, etc.), tape packing density, tape tracks, and the presence of standard (e.g., OS) "labels". In other words, considerations here must relate to compatibility or ease of convertibility to formats suitable for processing with NASIC hardware and available software, including the amount of pre-processing necessary to integrate the handling of a particular data base with others that NASIC may be processing. A related consideration is the amount of documentation the data base producer is able and willing to supply and his willingness to provide timely information on anticipated changes to the contents and format of the files.

Sturdivant (26), reporting experience at Marathon Oil Co., has confirmed that some data bases will be more difficult to implement than others. He reports that NTIS files presented great difficulties in conversion to the common format adopted (API) because of "extremely complex coding". The files of the U.S. Geological Survey were also difficult to use. The GEO-REF files of the American Geological Institute were easiest to re-format.

In situations in which NASIC is able to acquire searching "software" from the supplier of a data base, the data base and the related software need to be evaluated together in terms of the searching capabilities that the software provides. This takes us beyond the characteristics of the data base as such and into considerations involving the mode of acquisition and implementation of a service based on a data base. This type of consideration will be dealt with in the next section of this report.

D. ACQUISITION OF DATA BASES OR DATA BASE SERVICES: ALTERNATIVES

For any particular collection of data bases NASIC may offer services by any of the modes identified in Table 2. The major choice is the one between (a) acquiring a data base for in-house use, and (b) acquiring services from that data base from a service center already in existence. In actual fact, a completely free choice does not really exist, at least for several data bases of possible interest. In the first place NASIC, under the terms of its agreement with NSF, has a commitment to exploit the use of existing academic information centers as much as possible. Moreover, where such services exist for a particular data base and are well established, it is unlikely that NASIC could compete economically with these services (i.e., in most cases it is unlikely that an in-house operation could offer services more cheaply than those available through existing service centers). However, because of the broad scope of its activities, NASIC will undoubtedly identify some data bases for use that are not handled by existing information centers. For these data bases NASIC may elect to offer services directly, through acquisition of the data base for in-house searching. For this reason, then, it is worth giving some consideration to the entire range of "acquisition" possibilities.

Although a few data bases are available on a straight purchase or subscription basis, the majority of machine-readable files must be leased or licensed. Generally, a leas-

TABLE 2

MEANS BY WHICH DATA BASES OR DATA BASE SERVICES CAN BE ACQUIRED

1. NASIC obtains data base by purchase or licensing arrangement for in-house manipulation.
   a. Off-line
   b. On-line
2. NASIC purchases services from another agency:
   a. From the producer of the data base
      i. Off-line service
      ii. On-line service
   b. From an existing service center (retailer)
      i. Off-line service
      ii. On-line service

ing arrangement allows the recipient organization to offer services to its own staff members only. For example, a particular industrial organization may lease a machine-readable bibliographic file in order to offer retrospective search and/or SDI services within the company. If, as is the case with NASIC, the recipient organization is to offer services to a wider community of users, on a fee basis, it must enter into a licensing agreement with the supplier of the data base.

Although leasing and licensing arrangements and conditions vary somewhat from one data base producer to another, there is some commonality between them. In general, a data base is leased at a fixed annual fee. A licensing arrangement on the other hand, is based on a fixed annual fee plus some form of royalty arrangement related to amount of use. As an illustration we can consider the situation pertaining to the use of the COMPENDEX Tapes of Engineering Index Inc. The current files can be licensed for an annual fee of $6000. In addition, the licensee pays to Engineering Index $2.50 per year for each SDI profile it runs, up to the first hundred such profiles. After the first hundred profiles a sliding scale comes into effect, reducing to $1.70 per profile per year for each profile over 1000 serviced. For retrospective search purposes the arrangement is somewhat different. Four years of the data base, 1969-1972, can be licensed at an annual base fee of $21,600. The royalty for retrospective services is $2.00 per query per year of COMPENDEX searched. Alternatively, the licensee may pay a flat annual fee of $500 per year of COMPENDEX searched, with no limit imposed on the number of searches conducted. Where COMPENDEX is made available for on-line searching (as it is, for example, through Lehigh University) the royalty claimed by Engineering Index is 10% of the charge paid by the user for each search.

In the case of COMPENDEX, then, the royalty arrangement is based upon the number of searches conducted, or profiles services, per year. Other data base suppliers may have a somewhat different basis for royalty assessment. The International Food Information Service (IFIS) royalties, for example, are based on the number of citations supplied to customers.

Some additional points relating to leasing or licensing arrangements are also worth noting:

1. Some suppliers of data bases, including the Institute for Scientific Information, offer a year of "grace" as far as royalties are concerned. That is, royalties for outside use of a data base are imposed only after the first year of operation.
2. Sometimes the subscription to a tape service is tied to a subscription to the equivalent printed index. That is, an organization can subscribe to the magnetic tapes only if it also subscribes to the printed index. If the subscription to the latter is cancelled, the tapes must be returned to the supplier. This is the situation, for example, in the case of tapes available from the American Society for Metals and from the International Food Information Service.

3. Postage and cost of magnetic tape reels is normally charged in addition to the leasing or licensing base fee.

4. Cost of leasing or licensing may vary slightly, depending upon how frequently the lessee (licensee) requires his file to be updated. He pays less if he accepts less frequent updates of his file.

5. A leasing or licensing arrangement allows an information center to acquire a machine-readable data base and that only. Such an arrangement does not also provide searching software. If the data base supplier has such software available, this must be acquired by a separate purchase. In actual fact, many data base suppliers do not have searching software available. The lessee (licensee) must develop his own software or must acquire existing software from elsewhere (i.e., from another service center). The ASIDIC Survey of Information Center Services (27) lists 24 centers making software available to customers.

6. It is unlikely that NASIC could obtain exclusive rights to provide service from a particular data base for the Northeast United States, although it is possible that for certain data bases it might be possible to obtain exclusive rights for the whole country (in the way that the United Kingdom Chemical Information Service holds exclusive rights in that country for the products of the Chemical Abstracts Service), especially in the case of data bases of non-U.S. origin.

Let us now turn our attention to the acquisition of bibliographic services from an existing supplier. Possible suppliers to be considered are enumerated in Table 3. For some data bases, only one avenue of access is open to NASIC. Service from the New York Times Information Bank, for instance, can be achieved only through a subscription arrangement allowing remote on-line access to the files maintained by the producer. In other cases, however, a number of possibilities for service exist. In the case of COMPENDEX, for example. the tapes can be obtained under a licensing arrangement for in-house processing, batch processing services can be purchased through the University of Georgia or the University of Pittsburgh (to name only two), or the data base could be accessed remotely through the UCLA batch and on-line system through ARPA. In the case of some data bases, the choice may be one between the purchase of services directly from the producer (wholesaler) or the purchase of services from a middleman (retailer); some may involve a choice between the licensing of the data base or the purchase of service from the supplier of the data base; some may merely involve a choice between different middlemen (retailers).

In theory all of these options are open to NASIC, although in practice established policy constraints may limit the options available. Nevertheless, for certain data bases at least, NASIC will be faced with various service options and the general possibilities should all be examined for this reason. Several major producers of data bases do not themselves offer services based upon these files. Notable examples are Chemical Abstracts Ser-

TABLE 3

POSSIBLE SUPPLIERS OF BIBLIOGRAPHIC SERVICES

1. A data base producer offering batch processing services, e.g., BIOSIS.
2. The academic information centers, founded with NSF support, offering batch processing services, e.g., University of Georgia.
3. Other scientific information dissemination centers offering batch processing services, e.g., IITRI.
4. A data base producer offering direct on-line access, e.g., The New York Times.
5. An academic information center offering on-line access, e.g., Stanford.
6. Other licensees offering on-line access to data bases, e.g., Lockheed, System Development Corporation.

vice and Engineering Index Inc. The choice here is between acquiring the data base through a licensing arrangement or purchasing service from an existing service center. Where service from a particular data base is already available from a service center, this mode of access will normally be the preferred one. NASIC is expected to capitalize on existing centers as much as possible. Moreover, it is unlikely that NASIC would offer services more economically as an in-house operation, unless the volume of searches conducted by NASIC was large enough to reduce the unit cost per search to quite a small figure. Even in this situation, however, in-house operation is not necessarily more economical because NASIC could undoubtedly negotiate with an existing service center to achieve greatly reduced rates for a large volume of business. It is not possible to lay down any hard and fast figures here. For any particular data base in which this choice is involved NASIC would need to:

1. Estimate annual demand on the data base (both in terms of retrospective searches and SDI profiles);
2. Calculate costs of licensing the data base, acquiring the software necessary to search it (e.g., through purchase from another service center), and operating the data base on equipment available to NASIC.
3. Divide these costs by the projected demand in order to arrive at an estimated unit cost per retrospective search and an estimated unit cost per SDI profile per year.
4. Enter into negotiations with existing suppliers in order

to obtain a unit cost quotation for the estimated annual volume of NASIC business.

5. Compare the most favorable costs from these service centers with estimated unit costs for in-house operation.

In certain cases a choice may exist between purchase of service from a service center and purchase of service directly from the producer of the data base. Here cost comparisons must again be made. It is not unlikely that in this situation the most cost-effective approach would be to deal directly with the data base producer. For example, BIOSIS offers an SDI service, Current Literature Alerting Service (CLASS), based on tapes of BA Previews at a cost of $120 per annum per profile. This service purchased through the North Carolina Science and Technology Research Center costs $200 per annum; purchased through IITRI, it costs $250 per annum. These figures were derived from the ASIDIC Survey of Information Center Services, compiled by Williams and Stewart (27).

For some data bases, especially those that are relatively small and highly specialized, the only method by which NASIC can offer service may be to acquire the data base (by purchase or license) and provide service directly from the NASIC center (i.e., situations in which the supplier of the data base does not offer service and no service on this data base is now offered by existing service centers). In other cases, no service on a particular file is available through a service center and the choice lies between acquisition of the data base for in-house use or purchase of service directly from the producer. This choice must again be based on cost estimates and it will be volume-dependent. As an example, consider the Comprehensive Data Base of Patents (Chemical) available from IFI/Plenum Data Corp. This data base, of about 283,000 chemical patents going back to 1950, can be purchased (together with appropriate searching programs) for $34,000, plus $16,000 per annum for each yearly update. Over a five year period, with four annual updates, this data base would cost $99,000 to acquire, or about $20,000 per annum, exclusive of in-house operating costs for loading the data base and running searches. Only with a volume of retrospective search demand well in excess of 200 searches per year would it be likely that NASIC could operate this data base in-house at a lower unit cost per search than it could purchase service from the IFI/Plenum service bureau ($150 per search).

On the other hand, some of the smaller specialized data files may be operated in-house by NASIC more cheaply on even a relatively small volume of demand. For example, the bibliographic tapes of the Crystallographic Data Centre (University Chemical Laboratory, Cambridge, England) and the searching programs to interrogate them can be acquired for an initial cost of about $3000 and annual update costs in the region of $200. Over a five year period comparatively few searches may be needed to make in-house operation less expensive than the purchase of service from Cambridge at a minimum cost of around $25 per search.

It seems likely that NASIC will operate in a number of different modes. In fact, when fully operational, all the access modes of Table 2 may be used. The majority of data bases are likely to be accessed though existing service organizations, in both on-line and off-line modes, some others will be accessed through the data base producer (both on-line and off-line), and a few may be processed in-house by NASIC itself. One problem that NASIC must face is the problem of choosing between various service centers, where a particular data base is available through several such centers. Some guidelines relating to this choice are given in the next section of this report.

E. GUIDELINES ON THE CHOICE OF SERVICE CENTERS

Given that a number of different suppliers of services exist for a particular data base (SDI service based on Chemical Condensates is, for example, available from ARAC, IITRI, the University of Pittsburgh, UCLA, the University of Georgia, the North Carolina Science and Technology Research Center, the American Petroleum Institute, and the Institute of Paper Chemistry), what guidelines can be used by NASIC to help in deciding which information center to use? For certain data bases an important consideration will be the decision as to whether to acquire access on a batch processing basis or in an on-line mode. In the following paragraphs some considerations relating to the choice of service center are discussed. The fact that, as a matter of policy, NASIC may wish to give preference to the six academic information centers previously mentioned is tacitly assumed and will not be considered directly in the discussion.

Before we can discuss the comparison or evaluation of various service centers, however, it is necessary for us first to consider by what criteria information services are evaluated by their users. The most important of these criteria are enumerated in Table 4. In general, any service or product is judged in terms of cost, time and quality factors. An information service is no different from other types of service in this respect. The service must be provided at a cost that the user feels is reasonable in relation to the benefits associated with it. Cost to the user involves more than direct charges. It in-

TABLE 4

CRITERIA BY WHICH USERS EVALUATE INFORMATION SERVICES

Cost

- Direct charges
- Effort involved in use
    - Ease of interrogating system
    - Form of output provided
    - Backup document delivery capability

Response Time

Quality Considerations

- Coverage (completeness)
- Recall
- Precision
- Novelty
- Accuracy of data

Cost-Effectiveness (a quality consideration)

- Cost per relevant document or reference supplied.
- Cost per new relevant document or reference supplied (i.e. novelty-cost ratio).

cludes the cost of his own time, i.e., how much effort is involved in use of the system. Studies of the information seeking behavior of scientists and other professionals have consistently shown that accessibility and ease of use are the prime factors influencing choice of an information source. In general, the most convenient source of information will be chosen, whether or not this source is perceived by the user to be the most comprehensive, authoritative or, in some sense, the "best". Ease of use factors include ease of interrogating the system in the first place (i.e., ease of making one's needs known) and ease of use of the product provided by the system (i.e., the form of output supplied). A very important facet of the latter is the availability of an efficient and convenient document delivery capability. A service that delivers bibliographic citations goes only part of the way toward satisfying an individual's information needs. Such a service causes considerable frustration if the user is unable to obtain the documents cited or can only do so through procedures that he views as inconvenient and time-consuming.

The users of information services may be viewed as having four major types of information needs:

1. Specific factual information of the type that might come from some type of reference book or from a machine-readable data bank (e.g., thermophysical property data on a particular substance).
2. A few "good" articles (or references to them) on a specific topic.
3. A comprehensive literature search in a particular subject area.

4. A current alerting service whereby the user is kept informed of new literature relevant to his current professional interests.

These different needs have different response time requirements associated with them. The requirement relating to the current alerting service is that it should deliver regularly and frequently and that the information supplied should be as current as possible. The user needing a comprehensive literature search is usually a person engaged in a relatively long-term research project. Speed of response is usually not critical to him (except that there may be some date beyond which the search results will have no value or, at least, greatly reduced value); he is willing to wait longer in order to achieve completeness (i.e., completeness is more important to him than speed). For the other types of information needs, on the other hand, the user will generally want rapid response, in fact the type of turnaround time that is usually associated only with on-line systems.

Beyond cost, effort and time factors, the user will also be concerned with the quality of the product provided. The quality considerations of most concern to him are the completeness of the data base (i.e., <u>coverage</u>), the completeness of a particular search (i.e., <u>recall</u>), the degree of relevance of the search results (i.e., <u>precision</u>), the <u>novelty</u> of the results (important in the case of a current awareness service, which is really only valuable if it brings things to a user's attention before he learns of them by other means), and the <u>accuracy</u> of the results (which is a quality factor related to data retrieval systems rather than bibliographic systems).

All of these user criteria must be borne in mind by NASIC in planning its services and operations. It is clear that these criteria influence many decisions that must be taken. Some of them are related primarily to data base characteristics and influence choice of data base, as discussed earlier in this report. Other criteria relate to NASIC's own organization and facilities (e.g., ease of use, document delivery backup), while yet other performance factors will influence choice of centers providing service to NASIC. It is these that we are primarily concerned with at this point.

In choosing service centers NASIC must seek to identify those delivering the highest quality of product with the least processing delay and at the least cost. Unfortunately, these requirements tend to be conflicting. We must usually pay a higher price for quality and we may have to wait longer to achieve it.

A major decision that NASIC must face for several data bases is the decision as to whether to access the data base through a center offering off-line, batch processing services or through a service organization (e.g., Lockheed, SDC) making the data base available for direct on-line interrogation from remote sites. For some data bases there is no choice available. The New York Times Information Bank, for example, can only be accessed on-line, and the off-line MEDLARS operations are now almost entirely phased out in favor of MEDLINE. To return to the four types of information need noted earlier, the current alerting activity is one which is probably handled most effectively by off-line, batch processing (except that the construction of user interest profiles is probably best handled by heuristic interaction with an on-line system). The comprehensive literature search is also best handled in the off-line mode, where no stringent deadlines are imposed by the user, although again there are obvious advantages associated with being able to test out a strategy on-line on part of a data base before committing this strategy to the more expensive search of the complete file. For the other types of information needs, however, the response time requirements of the user are likely to be such that only rapid, direct access to the on-line data base will be acceptable. It seems, then, that NASIC must consider both off-line and on-line suppliers of service and that for some data bases both modes of searching must be made available. NASIC will, then, at some point in time, find itself in the position of evaluating on-line service centers as well as off-line service centers.

Parenthetically it is worth pointing out that on-line access to a data base, through a service center, may well be cheaper for a large volume of demand than off-line retrospective searches, as well as offering more rapid response and the possibility of improved results through the interactive nature of the search. To illustrate this, consider the services offered by Lockheed Information Retrieval Services, as depicted in Table 5. Assume the highest level of fixed costs, namely $720 a month ($420 a month for rental of high speed display/printer plus $300 a month for communications costs), that the capital outlay on installation and service is amortized over two years

TABLE 5

ON-LINE INFORMATION SERVICE AVAILABLE FROM LOCKHEED INFORMATION RETRIEVAL SERVICES

DATA BASES AVAILABLE

National Technical Information Service (NTIS)

ERIC - Research in Education (RIE)

ERIC - Current Index to Journals in Education (CIJE)

Exceptional Children Abstracts

PANDEX

COSTS

1.	Installation and service (one-time cost)	$500 to $1000
2.	Rental of high speed display/printer terminal operating at 240 CPS	$420/month
3.	Communications costs	$200-300/month
4.	Computer charges	$25-35/hour
5.	Off-line printing cost	10¢ per item (citation or abstract)
6.	Nominal royalty charge (for certain data bases only)	

for about $30 a month, and that the average search takes 15 - 20 minutes on-line and produces a printout of 50 citations or abstracts. Assume a volume of demand, over all five data bases, of fifty searches a month (not an unreasonable level of demand for such data bases within the Northeast academic community as a whole). The cost per search would then be about $30 ($750/50 = $15 plus $10 per search computer costs plus $5 per search printout costs). This compares with a figure quoted by the University of Georgia of $35 <u>per volume</u> for a retrospective off-line search of RIE and $35 <u>per volume</u> for a retrospective search of CIJE. The same figure is quoted for a search through the Government Reports Announcements (NTIS) files.

Remember also that these results put the on-line service in the worst possible light as far as costs are concerned because the full cost of the terminal rental and communications is charged to the use of only five data bases. If Lockheed adds additional data bases (or if the terminal is also used to access other data bases elsewhere) and the cost is then spread over a larger volume of usage, the cost per search for on-line access could be reduced considerably. With ten data bases accessible from the same terminal, and with a volume of use on all ten data bases of 200 searches/month, the cost per search could be $20 or less. (Tables describing the main features of a number of available on-line systems are presented in Appendix 2 of the Attachment).

Let us now return to the user-oriented evaluation criteria of Table 4 and give some further consideration to their implications for the selection of service centers. Since NASIC is to be a self-supporting operation, cost considerations will be of major importance. Centers must be evaluated in terms of the cost of their services and because costs of information services tend to be volume-dependent, NASIC must endeavor to negotiate the best possible terms for the projected volume of business it will generate on a particular data base. For retrospective search purposes the cost of dealing with a center operating in the off-line mode must be compared with the cost of accessing data bases from terminals located within a NASIC center. In fact, this cost analysis may well be the most important one. The difference in charges among the various off-line centers is usually not very great, at least for several of the data bases.

These charges are laid out conveniently in the ASIDIC Survey (27) of 1972, although these data are now a little out-of-date. As an example, the North Carolina center quotes a cost for SDI service on the ERIC tapes of $75 per profile per year. The Georgia center quotes a cost of $80 for the same service. On the other hand, for an SDI service based on CA Condensates, IITRI quotes a flat rate of $250 per profile per year, while Georgia quotes $260 for academic customers and $364 for commercial customers. In contrast, however, the Aerospace Research Applications Center (ARAC) quotes a charge of only $195 for this same service.

The response time requirement relates only to retrospective searches. In assessing this feature of the various centers, NASIC will need to contact a representative group of existing customers of these organizations. The <u>ASIDIC Survey</u> indicates that some processing centers operate with a short turnaround time in the range of 1 - 5 days, while others offer a service much less satisfactory, up to 20 days in some cases. Unfortunately, these data are presented in gross form only, the individual centers not being identified in the ASIDIC tabulation. In the O'Donohue (22) survey, seven centers were compared on processing time for retrospective searches. Only three of these centers routinely processed searches in a time which O'Donohue regards as "prompt", namely less than two weeks. While analyses of this type are difficult to find in published form, other customers undoubtedly have made their own comparisons and data of this type may be made available to NASIC.

Somewhat related to the response time for retrospective searches is the time it takes a service center to get an SDI profile "up and running". O'Donohue (22) quotes a range of from 2 weeks to 6½ weeks (from initial inquiry to first machine printout) for five processing centers from which SDI service was received.

Before leaving the subject of response time it is worth noting that NASIC must endeavor to identify centers that are willing and able to handle certain requests in a "special processing" mode. That is, they should be capable of handling a special request on a "rush" basis where necessary.

While it is easy to compare centers in terms of their charges, and relatively easy to compare them from the viewpoint of response time, it is not at all easy to make this comparison in terms of the quality of the product provided, the major qualitative considerations that are at least partly controllable by a service center being the recall and precision of search results (see Table 4). This statement needs qualification, however. It is relatively easy to judge the performance of a center, at least in terms of the precision of its searching, through a period of experience with this center. It is difficult, however, to compare centers in terms of what their service is likely to be (i.e., before a contract is actually initiated). However, because NASIC is likely to be a very substantial customer it seems reasonable that it should negotiate a trial period for a number of profiles, with one or more service centers, before a formal subscription is placed with a center. In fact, NASIC might consider developing a small group of "test searches" (i.e., searches for which a known set of relevant documents is identified within a particular data base) and use these test searches to evaluate various processing centers in terms of both the recall and precision of the search results, as well as response times. O'Donohue (22) quotes some precision figures ("percent relevant") for SDI service from several centers and based on several data bases. These range from a low of 4% precision (the Georgia center searching the Nuclear Science Abstracts Tapes) to a high of 54% (IITRI searching CA Condensates Tapes). It should be noted, however, that while there is likely to be a certain minimum level of precision that is acceptable to a particular user, an extremely high level of precision would suggest that the profile is missing many of the relevant documents (i.e., recall is low) because there tends to be an inverse relationship between recall and precision in searching. For example, if a user interest profile consistently operates at 80% precision we can be almost certain that it is also running at a very low rate of recall. For both the current awareness need and the "comprehensive search" need, high recall is likely to be more important to the user than high precision (although precision below a certain level may be intolerable). It must also be recognized, however, that while a user is able to judge the precision of a search (i.e., determine what proportion of all citations delivered are relevant to his interests), he is usually in no position to be able to judge its recall because he does not know what the search may have missed. The recall ratio achieved for a particular SDI profile or a particular retrospective search can usually only be estimated by a specially devised test and analysis.

If we return now to a further examination of Figure 2, in which the entire range of factors governing recall and precision of a delegated search are displayed, it can readily be seen that several of these factors are associated with characteristics of the data base itself (i.e., indexing and vocabulary characteristics) and are essentially outside the control of the service center processing the data base. However, two very important factors influencing performance <u>are</u> within the control of the center, namely the quality of the interaction with the user (procedures by which his information need is "negotiated" with the system) and the quality of the searching strategies used, whether for SDI or retrospective search.

These performance factors are closely related to the precise modes of operation adopted by NASIC. Three broad modes of operation appear possible:

1. The scientist to be served is put into contact with the service center. Staff at this service center clarify his needs and prepare the user interest profile or strategy for a retrospective search.
2. The scientist discusses his need with a NASIC representative (e.g., an Information Services Librarian) who then relays his interpretation of this need to the service center where a search strategy or profile is constructed.
3. The scientist discusses his need with a NASIC representative who converts this need into a searching strategy or user interest profile that is then run at a service center.

In general, the second of these alternatives is the least desirable since the more intermediaries placed between a user and a data base, the less successful the search is likely to be. It is well known that when a message is relayed through a chain of people the possibility of distortion ("noise" in a communication sense) exists at each step in this chain. The first alternative is likely to produce the best results initially because of the experience that personnel at a service center will have accumulated in use of a particular data base. However, this mode of operation provides no training possibilities for NASIC staff. In the long run, the third alternative may be the best mode to adopt. Once NASIC staff have been trained to construct searching strategies or user interest profiles for a particular data base and set of searching programs, the fact that these staff members are "closer" to the user community (physically if nothing else) may result in improved user-system interaction and improved information products as a result.

Of course, this presupposes that the service center dealt with will allow NASIC to operate in this way. If the third of the alternatives is the one that is organizationally most acceptable to NASIC, an important selection criterion will obviously be whether or not the service center will permit a mode of operation in which search strategies or interest profiles are prepared by NASIC staff and simply "run" (in a machine sense) at the center. This mode of operation also presupposes that the center has available adequate facilities and materials for training people in search techniques. A center that has no training program and no adequate search manuals (which should be data-base related) will be unacceptable to NASIC if the third of the processing alternatives is the preferred one. It should be pointed out that some processing centers have produced very excellent searching guides and searching tools, both of a general nature and related to particular data bases. An example is IITRI, which has produced an excellent Search Manual, with supplemental guides for particular data bases, as well as a very useful word truncation guide. The University of Georgia has likewise prepared a very complete Profile Coding and Management Manual, conducts regular workshops in profile construction techniques, and provides on-line access to representative subsets of various data bases for training purposes. This training program has been described by Park et al (28).

As far as the on-line approach to data bases is concerned, one assumption is that NASIC personnel will search these data bases for users by means of terminals located in a NASIC center. In this case it is clearly imperative that NASIC staff or representatives (ISL's) be well trained in the search procedures associated with a particular system and the search techniques needed to exploit a particular data base effectively. If NASIC deals with an on-line service center to gain access to various data bases, this center should be capable of providing the necessary training, as well as appropriate searching aids.

As previously mentioned, the factors affecting performance of an information service that are not primarily input-related (indexing and vocabulary factors), and thus outside the direct control of NASIC or some service center, relate to interaction with the user and the quality of searching strategies. In other words, the quality of the service will be heavily dependent upon the quality of the information staff who are interacting with the users and the quality of the information staff preparing search strategies or interest profiles (whether these be NASIC staff members or personnel associated with a processing center), the training of these staff members, and their degree of experience with particular data bases and particular searching software.

Clearly, however, we have just identified another important variable affecting the performance of an information service, and thus the choice of a processing center, namely the capabilities of the searching software. In general, different service centers, both those operating on-line and those operating off-line, are using different searching software. The IITRI searching programs are different from those used in Georgia, the Lockheed programs (DIALOG) are different from those used by SDC (ORBIT), and all of these are entirely different from the LEADER system of Lehigh.

In assessing the capabilities of a processing center, NASIC must assess the capabilities of the search programs in use by that center as well as the output options available from these programs. This will be particularly critical in relation to data bases that NASIC staff members interrogate directly themselves (on-line or off-line). While all such programs have the same general objectives and capabilities, there are differences between them at the specific feature level. The off-line search programs, particularly those searching text, must go beyond simple Boolean AND, OR and NOT capabilities. Weighted term searching (permitting the ranking of output) is an important capability and nested search logic is an essential requirement for systems operating exclusively in a Boolean search mode. For text searching, word truncation capabilities (both left and right truncation) are essential and word proximity operators (i.e., the ability to specify how close two words should be in text before they are considered to be related) are highly desirable. A number of output formats should also be available, both in terms of what is printed (citation, abstract, etc.), in what sequence it is printed (i.e., sorting options) and on what it is printed (i.e., the output medium).

In the evaluation of on-line searching systems additional requirements become important. These requirements include the capability of developing term lists, thesauri, or other searching aids, and the capability of providing various tutorial and "help" features to the searcher.

Fortunately for NASIC, much of the work on the comparison of searching software, as well as output capabilities, has already been done. The ASIDIC Survey (27) has compared the search and output features of the off-line processing centers, including full members, associate members and non-members of ASIDIC. These data are presented in gross form in Table 6 - 8 and in a more exact form (i.e., each feature associated with each center) in Tables 9 - 11.

The comparison of on-line searching systems has been done by Stanford University, Institute for Communication Research, under a grant from the National Science Foundation. The features of the major on-line searching systems were discussed and tabulated at an important meeting that took place at Stanford in April, 1973. One of the authors of the present report, F. W. Lancaster, attended this meeting and serves as an advisor to Stanford on this project. As a result of the meeting he prepared a full report entitled <u>The Present Status of On-Line Interactive Retrieval Systems</u>. This report contains a complete summary of the searching, output, training, tutorial and monitoring features of the major extant on-line searching systems. Because of its potential value to NASIC it is included intact as Attachment 1 of this report.

Before leaving this subject of service centers and their selection, some additional comments need to be made. Selection of a service center will also involve considerations of experience,

TABLE 6

CURRENT AWARENESS SEARCH SOFTWARE CAPABILITIES

System Feature	Number of Information Centers: Full Members	Associate Members	Non-Members	Total
Weighted profile terms	16	7	3	26
Truncation:				
Left	13	8	6	27
Right	22	9	10	41
Both simultaneously	13	8	6	27
No truncation	3	8	3	14
AND, OR, & NOT logic operators	22	14	12	48
Proximity logic operator	6	3	1	10
Nested logic ((( )))	13	11	9	33

TABLE 7

CONTENT OF CURRENT AWARENESS OUTPUT

Information Printed	Number of Information Centers: Full Members	Associate Members	Non-Members	Total
Abstracts	20	6	10	36
Keyword(s) and/or index term(s)	20	12	9	41
Term(s) causing hit	12	8	5	25
User name with each citation	9	7	3	19
User ID number with each citation	19	8	5	32

TABLE 8

MEDIA FOR CURRENT AWARENESS OUTPUT

Type of Output Available	Full Members	Associate Members	Non-Members	Total
Cards; 1-part:				
3" x 5"	0	0	1	1
3-¼" x 7-3/8"	2	2	0	4
4" x 6"	3	1	2	6
4" x 9"	0	0	1	1
5" x 8"	2	3	0	5
Cards; 2-part:				
3" x 5"	1	0	0	1
3-¼" x 7-3/8"	9	0	0	9
4" x 6"	0	0	0	0
4" x 9"	0	0	1	1
5" x 8"	0	0	0	0
Paper Listing:				
2-½" x 7-½"	1	0	0	1
7-½" x 11"	0	1	0	1
8-½" x 11"	12	6	9	27
11" x 14"	5	11	6	22
14" x 17"	0	1	0	1
Multiple copies available	12	10	7	29
COM	6	0	1	7
Tape	8	2	1	11
Multilith masters	2	1	2	5

TABLE 9 (cont.)

	ARAC	B&W	CAN/SDI	COSTI	DOW	DUPONT	GE	GTE	IBM	IITRI	KAS	KODAK	LEHIGH	LIVERMORE	NTIS	N. CAROLINA	OSU	31 CO.	3M--PAT. & PROF.	3M--TECH. COMM.	U. CALGARY	UCLA	UGACC	U. KANSAS	WRISC	XEROX
Service available to:																										
Public	●		●	●						●	●		●		●	●	●	●			●	●	●		●	
In-house Only		●			●	●	●	●	●			●		●					●	●				●		●
Output format:																										
Cards (1- or 2-part):																										
3" x 5"																	2									
3-1/4" x 7-3/8"		2	2	2				1			2							1&2		2	2					2
4" x 6"	1				1							2											1			
4" x 9"																										
5" x 8"										1								1								
Paper:																										
3-1/2" x 7-1/2"									●																	
7-1/4" x 11"																										
8-1/2" x 11"	●	●				●	●				●		●		●	●		●	●				●		●	
11" x 14"												●	●	●								●		●		
14" x 17"																										
Output content:																										
Abstract printed		●	●	●	●	●		●	●	●		●	●	●	●	●		●	●	●	●	●		●	●	●
Keywords printed	●	●	●	●	●	●				●	●	●	●	●			●	●		●	●	●	●	●	●	●
Hit terms printed	●	●	●	●	●		●			●		●	●	●			●								●	
SDI search software:																										
Weighted terms	●		●		●	●	●			●	●		●				●	●	●	●		●	●	●		●
Left truncation			●		●	●	●			●	●	●		●				●	●			●	●	●		
Right truncation	●	●	●	●	●	●	●	●	●	●	●	●	●	●			●	●	●		●	●	●	●		●
Left & Right truncation			●		●	●	●			●	●	●		●				●	●			●	●	●		
AND, OR, NOT logic		●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	
Proximity logic		●						●	●				●								●	●				
Nested logic ((()))			●		●		●			●	●	●	●	●		●						●	●	●	●	
Data base reformat	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●		●	●	●	●	●	●

SUMMARY OF INFORMATION CENTER SEARCH AND OUTPUT CHARACTERISTICS

ASIDIC FULL MEMBERS

TABLE 10

	API	ASM	BA	CID/ENDS	CTC	GEO-REF	IOWA	IST	JOHNS HOPKINS	KAROLINSKA	MIT	NORTHWESTERN U.	ORBIT/SIS	SIE	SXE	U. LOUISVILLE	U. NEBRASKA	U. VIRGINIA
[illegible]																		
Books	●	●	●	●	●	●	●	●		●			●	●		●		
In-house [illegible]									●		●	●			●		●	●
Output [illegible]																		
Cards [illegible]																		
3" x 5"																		
3-1/4" x 7-3/8"	1						1											
4" x 6"				1														
4" x 9"																		
5" x 8"			1		1								1					
Paper:																		
3-1/2" x 7-1/2"																		
7-1/4" x 11"								●										
8-1/2" x 11"	●	●					●		●		●		●					
11" x 14"			●		●	●					●	●	●	●	●	●	●	●
14" x 17"		●																
Output content:																		
Abstract printed					●	●			●		●		●					
Keywords printed	●		●		●		●	●	●	●	●	●	●	●		●		
Hit terms printed	●				●	●	●	●			●					●		●
SDI search software:																		
Weighted terms	●		●		●			●					●			●		●
Left truncation	●		●					●		●	●		●		●	●		
Right truncation	●		●		●			●		●	●		●		●	●		
Left & Right truncation	●		●					●		●	●		●		●	●		
AND, OR, NOT logic	●	●	●	●	●		●	●	●	●	●		●		●	●	●	
Proximity logic									●		●						●	
Nested logic ((()))	●	●		●	●		●	●	●	●	●				●		●	
Data base reformat	●	●	●	●					●	●	●		●					

SUMMARY OF INFORMATION CENTER SEARCH AND OUTPUT CHARACTERISTICS

ASIDIC ASSOCIATE MEMBERS

TABLE 11

	ABBOTT	ARS/USDA	DCRT	EDA	FINLAND	HRIS	IPC	MRIS	NETHERLANDS	OXFORD	TNRIS	TRAIS	UKCIS	WESRAC
[illegible]					●		●	●	●		●		●	●
[illegible]	●	●	●	●		●				●		●		
[illegible]														1
[illegible]														
[illegible]									1				1	
[illegible]					1&2									
[illegible]														
Paper, 3-1/2" x [illegible]-1/2"														
[illegible]-1/4" x 11"														
8-1/2" x 11"	●	●				●	●	●			●	●	●	●
11" x 15"		●	●	●				●			●	●		
14" x 17"														
Output content:														
Abstract printed		●	●	●		●	●	●	●		●	●		●
Keywords printed		●	●	●	●	●	●	●	●					●
Hit terms printed			●		●	●	●						●	
SDI search software:														
Weighted terms		●		●					●					
Left truncation		●	●	●			●		●				●	
Right truncation		●	●	●	●		●	●	●		●	●	●	
Left & Right truncation		●	●	●			●		●				●	
AND, OR, NOT logic	●	●	●	●	●	●	●	●	●		●	●	●	●
Proximity logic													●	
Nested logic ((()))		●	●	●	●	●	●	●			●	●		
Data base reformat		●	●		●		●						●	●

SUMMARY OF INFORMATION CENTER SEARCH AND OUTPUT CHARACTERISTICS

ASIDIC NON MEMBERS

proven reliability, flexibility of operation (e.g., the ability to accommodate high priority searches on a "rush"basis), as well as general attitude and "customer orientation" (the "personal element" in O'Donohue's evaluation). The experience of the information center is important, especially its experience with particular data bases. Experience is related to years of operation and to volume of profiles and/or searches handled. These data are readily gleaned from the ASIDIC Survey. Clearly, NASIC should deal only with centers that appear to be stable and whose continuity appears to be assured. Another element to be considered is the degree of interest that the center exhibits toward the quality control and improvement of its products. One form of evidence of this is the amount of interaction and iteration a center will undertake before it "stabilizes" a profile. Another form of evidence is the amount and type of feedback and evaluation solicited from users, and the degree to which the center attempts to improve its performance on the basis of such feedback and evaluation. Some gross information on solicitation of feedback is presented in the ASIDIC Survey. The degree to which user citation is used as a means of improving overall performance is also evidence of a center's interest in the quality of its product. O'Donohue's survey is a useful summary of experience with a small number of centers. It is important to note that some centers satisfied the evaluation criteria of this study very well while one or two others were highly unsatisfactory. O'Donohue concluded that "great care is required in the selection of commercial information services. The spectrum of potential satisfaction is wide and the user must analyze his needs and his suppliers' capabilities carefully to optimize results".

Ultimately a user will judge an information service on cost-effectiveness grounds, relating the cost of using the service to the quality of the product provided by that service. The most useful cost-effectiveness measure to use in the evaluation of information services is the cost per relevant citation obtained from the service. If a user subscribes to an SDI service, at an annual cost of $150, and is supplied with seventy-five relevant citations in a particular year, the cost per relevant citation is $2.00. When fully operational NASIC must develop quality control and monitoring operations that will permit the cost-effectiveness evaluation of data bases, and services from them, in terms of this important measure. This will require the development of procedures for obtaining regular and precise feedback from users. Perhaps the most complete evaluation of data bases and services yet conducted, although restricted to drug-related information, is that reported by Ashmole et al (25). These investigators compared various approaches to locating information on a particular drug in terms of the yield of each source, the number of unique references supplied by each source, the source that disclosed a particular reference for the first time (novelty), and cost per relevant citation.

In conclusion, it seems likely that the NASIC operation, when fully implemented, will involve a number of service modes. Some data bases NASIC may choose to acquire and search in-house, on-line or off-line, while others will be accessed through service centers. It is likely that a number of service centers will be used for different data bases and that, to serve different purposes, the same data base may be accessed both through an on-line processing center and an off-line center. This section of the report has attempted to present various guidelines and criteria to guide NASIC in the choice of centers with which to deal. Although machine-readable data bases have been assumed throughout this discussion, it must be remembered that some unique data bases exist in manual form in various parts of the world and that both SDI and retrospective search services are available from centers without mechanized operation. An important example is the Scientific Documentation Centre, Dunfermline, Scotland, which operates a unique service in the field of spectra and spectral data. It is important that NASIC seek out such centers and, where appropriate, integrate these services into the overall NASIC operation.

F. SOME CONSIDERATIONS RELATING TO NASIC PRODUCTS AND SERVICES

In general, as discussed earlier, information services are of two broad types:

1. Current awareness services (for alerting purposes) such as SDI.
2. Retrospective search services (on-demand).

Both types of service can be provided from a particular data base and both may be offered by a particular service center. Retrospective searches may themselves be divided into two types:

a. Quick-reference search to find a particular item of data or a "few good references" on a specific topic.
b. Comprehensive search to find all references on a particular subject, or all references published on this subject in a particular period.

The first of these requirements can usually be satisfied by a relatively short search but this usually requires the rapid response that is most likely to be satisfied through a search of printed tools or machine data bases available on-line. The second type of search is more time-consuming but is usually associated with a much less stringent response time requirement. This type of search is more suitable for off-line processing in large machine-readable data bases, although the strategy might be tested initially on a portion of the data base available on-line.

It seems likely that in its implementation schedule NASIC should give first priority to the SDI aspect of its service. There are a number of reasons for this. An SDI service is generally more commercially viable than a retrospective search service, it is likely to have a wider initial appeal, and it is a service for which volume of demand is easier to assess. In contrast, it tends to be quite difficult to estimate the likely pattern of demand for retrospective searches from a particular data base. Moreover, because an assured minimum annual volume can be determined for SDI service from a particular data base, it will be possible for NASIC to negotiate favorable rates from a processing center. Likewise, the costs of licensing a data base, purchasing software to search it, and offering service as an in-house operation, are much more likely to be recovered through a regular current awareness service than through a ret-trospective search service. The latter is only likely to be ec-onomically viable if there is a fairly heavy and constant level of demand for service. In the case of an in-house batch processing service, the volume of demand must be such that batches of specified minimum size can be run on a regular basis, say once a week.

In the case of SDI service, it seems desirable, for reasons discussed earlier, that NASIC information services librarians be trained in techniques of profile construction and be given responsibility for development and updating of profiles, whether a data base is operated in-house by NASIC or through a service center. For certain data bases used for SDI purposes, NASIC may also have on-line access. In this case, an initial user interest profile can be developed by direct interaction with the data base until it is refined to the point at which it can be submitted for regular batch processing.

The use of an on-line system directly for SDI purposes is also possible, a good example being NLM's SDILINE. On-line SDI may play some part in the overall pattern of NASIC operation, but the use of this type of service is likely to be restricted to certain large academic libraries within the NASIC community, these libraries having their own on-line access to various data bases.

Group SDI tends to be considerably cheaper than completely customized SDI to an individual. By "group SDI" we mean service, based on the same profile, to a group of users having common, homogeneous interests. Several service organizations offer some form of group SDI. A good example is the "macroprofiles" service of the United Kingdom Chemical Information Service. Another is the ASCATOPICS service of the Institute for Scientific Information. A service of this type is also offered by IITRI and CAN/SDI. Group SDI works on the basis of standard profiles in relatively broad subject areas. These profiles are established by the service center and advertized by that center. If a standard profile is a reasonable match with the interests of a particular scientist he may subscribe to it at a rate considerably less than the rate he would pay to have a profile custom-made to his interests. Because NASIC is to service such a large audience, it should be possible to identify several groups of scientists with relatively similar interests. Indeed, this should be a high priority task in the NASIC program. Even if many of the interest groups thus identified do not match one of the group SDI profiles already in existence, it should be possible to negotiate a very favorable rate for a new group profile with an appropriate service center.

Before leaving the subject of SDI, it is worth pointing out that keeping up-to-date with newly published literature is only one facet of a complete current awareness program. Another important facet involves finding out what is going on in current research (i.e., who is doing what in a particular subject area). This involves the searching of indexes to ongoing research projects, the most notable of course, being the files of the Science Information Exchange. NASIC must certainly make use of this important source and other specialized sources of information on ongoing research. Access to information on ongoing research must be an integral element in a complete information service.

In the area of retrospective searching a much greater variety of approaches is possible. A rather gross summary of some major possibilities is given in Table 12. It seems likely that, when fully operational, NASIC will maintain certain data bases in-house, for SDI, retrospective search and data retrieval functions. Others will be accessed by on-line terminals located at NASIC headquarters or at another NASIC center. Yet others will be searched through the batch processing facilities of a service center or a data base producer.

One of the major functions of the NASIC information services librarian (ISL) will be that of deciding, for any particular information need presented to him, which available data base should be queried. He should also be able to quote to the customer a cost for the search to be conducted. The ISL will clearly need to have a wide knowledge of available bibliographic and data re-

TABLE 12

POSSIBLE MODES OF ACCESS TO DATA BASES

FOR TYPES OF INSTITUTIONS AND VOLUME

OF USE

DATA BASES, BY VOLUME OF USAGE

NASIC SERVICE INSTITUTION	HIGH VOLUME	INTERMEDIATE VOLUME	OCCASIONAL USE
Large academic library	Direct on-line access to data base	Access via NASIC HQ, where data base is accessible on-line	Access via NASIC HQ, where data base is accessed off-line through producer or service center
Medium-sized library	Access through teletype or telephone to NASIC HQ, where data base accessible on-line	Access via NASIC HQ, where data base is accessible on-line	Access via NASIC HQ, where data base is accessed off-line through producer or service center
Small library	Access through NASIC HQ, via mail or telephone	Access via NASIC HQ, where data base is accessible on-line	Access via NASIC HQ, where data base is accessed off-line through producer or service center

sources, manual as well as mechanized. A very important NASIC tool will be a printed guide to available resources. Such a guide should be much more comprehensive than any presently available, and it must be indexed by specific subject areas. The guide should be capable of indicating to the ISL which sources are most likely to yield information of a particular type. For a request for information on nuclear instrumentation it should be capable of leading him directly to the International Nuclear Information System (INIS), the Information Service on Nuclear Science and Technology (UKAEA) and Nuclear Science Abstracts (USAEC) as most likely sources. The detailed description of the data bases, and services available from them, will help the ISL decide which source is most likely to be profitable for the specific topic under review. Likewise, the guide should be capable of leading to PESTDOC for a search on rodenticides, to CAIN for a search relating to forestry, and to the services of the Highway Research Information Service (HRIS) for information on the design of parking decks. This NASIC guide must be kept up-to-date and might therefore be issued in loose-leaf form. Each ISL would be encouraged to contribute to it by bringing to the attention of the compilers newly discovered sources, not previously indexed, for a particular subject area.

Through the ISL a NASIC customer should have access to any data base, manual or mechanized, that exists and is generally available. In practice, NASIC must identify those data bases likely to satisfy the greatest volume of demand and make these data bases most accessible to the scientific community. A high-usage data base would be made readily accessible through an on-line terminal connection established at NASIC headquarters. Some of the larger NASIC libraries may also have direct on-line connection to one or more data bases expected to be of major value to their own user group. For smaller libraries and less-used data bases a common mode of service will be a rapid-response on-line search requested by telephone call or teletype message to NASIC headquarters. It is likely that the NASIC headquarters will have facilities to access a number of important data bases remotely. Additional data bases may be maintained in-house for on-line or off-line access. For the less-used data base, however, NASIC will request service from a service center when the need for such service arises. It is desirable, therefore, that the NASIC guide should identify a preferred service center for each data base. Arrangements should be made with this center for rapid turnaround service if the need for fast response is critical for a particular request.

Initially, at least, it is likely that most searches will be conducted by NASIC information specialists or information specialists associated with other centers. However, the NASIC headquarters may have an interactive information facility that is open to members of the scientific community. If a scientist wishes to visit this facility, and use the on-line resources directly, he may be allowed to do so. Similarly, some of the larger academic libraries, having their own on-line facility, may allow scientists access to data bases directly, and will provide some training facilities to make this possible. Finally, if within the NASIC community a large academic department is identified as having a very high level of demand for a particular data base, arrangements might be made to provide terminal access within that department. If NASIC is able to distribute a number of terminal installations throughout the region, and each of these installations can be used to access a number of different data bases, the combined volume of business thus generated may be more than sufficient to make this level of accessibility justified, both in terms of convenience and economics.

Since NASIC will be dealing with data as well as with bibliographic retrieval, an important element in the overall NASIC program may be a referral service. The information services librarian, with the assistance of staff at NASIC headquarters, should be capable of referring a scientist to any likely source of information, whether this is a formal data base, an information analysis center, or an individual consultant. In this important referral function, NASIC may work closely with the National Referral Center for Science and Technology.

Finally, at a point in time when NASIC is fully operational and terminals are widely available in academic institutions throughout the area, NASIC should consider the possibility of initiating a program, using computer facilities at NASIC headquarters, or a NASIC center, for providing on-line support to the creation and exploitation of personal files. Through this service an individual scientist could purchase access to programs that would allow him to build files of data or references, add to them, delete from them, search them on-line, obtain machine listings, and so on. Personal files consitute a very important source of information for most scientists. Such files are not replaced by general information services, however good these services are. Indeed, formal information services, especially SDI operations, are used to feed personal files. Unfortunately, the conventional personal file tends to be a simple pigeonhole system, with very limited access capabilities. On-line systems designed to aid the researcher in the efficient exploitation of his own files (e.g., RIQS at Northwestern University and AUTONOTE at the University of Michigan) have proven very popular in the academic community. Such a system might be an important element in overall system planning. Ideally, through a terminal located in his own office or department, a scientist should be capable of accessing his own files, or at least indexes to them, as well as accessing any of several outside data bases made available to him through NASIC. From the same terminal, then, he is given convenient, rapid access to a very wide range of bibliographic and data resources. This type of facility may substantially alter his present information gathering habits, and he may well judge that the benefits available from such a service more than offset the costs involved in using it. With a very large operation these costs may not be excessive.

G. SOFTWARE ASPECTS OF INFORMATION SERVICE CENTER OPERATION

The operations and procedures necessary to provide the information services and products desired of NASIC will of course be largely computer based. Thus, it will be necessary to discuss the various types of computer programs or software needed by NASIC to perform its information services function, and the limitations of and problems associated with such software. Criteria for selecting and evaluating such software will be presented in order to assist NASIC in choosing among the possible alternative sources or methods for providing the desired information services and products.

NASIC may decide to purchase, lease, or license a certain data base and the appropriate software for its own in-house use (or for use on a nearby computer on which NASIC has purchased time), or NASIC may simply purchase certain information services from a remotely-located information center. A different collection of software considerations apply whether NASIC uses a certain data base in-house or purchases services from a remote service center. But, certain software considerations apply to all information retrieval software no matter which group actually operates it. In the following discussion the evaluation and selection criteria that apply to all information retrieval software will be presented. Then, the problems that arise and the additional evaluation and selection criteria that must be considered will be given for the situation when it becomes necessary or desirable to use software on a data base, computer, or computer system different from the one for which it was written. The latter will be the situation if NASIC decides to obtain a certain data base and appropriate software for in-house operation rather than simply purchasing information services from a remotely-located center.

For the operation of an information retrieval system the following four groups of computer routines are required: (1) Input/Output routines, (2) Search routines, (3) Data Management routines, and (4) Data Base Transformation routines.

Input routines are those computer routines which operate on the user's request for information. They can correct the form of the user's request by checking for errors in formating, punctuation, or spelling; or they can elucidate or expand the user's request by requiring more specificity or precision in his request statement, or by imposing on him or informing him of areas closely related to his area of interest. Examples of the latter class of input routines are thesaurus lookup routines. The inputting of the user's request for information is frequently done on-line, though it can be done off-line. Output routines are those which put the user's output onto the medium (cards, paper, or magnetic tape) and into the format and order that the user had indicated, and make provision for routing this output to him. Software for input/output operations will be discussed further in one of the following chapters.

Search routines are those parts of the computer programming system that identify (by some sort of matching operation) those pieces of, or citations from, the desired data base(s) that correspond most closely to the user's expressed information wants. Following the identification of the information desired by the user, certain portions of this information are extracted or reproduced and sent to the output routines for transmission to the ultimate user. Searches in information retrieval systems are of two main types: current awareness searches and retrospective searches. Software for search operations will be discussed further in a following chapter.

Data Management routines are those routines which handle the complex billing and accounting procedures required for each user of the information retrieval system and which also perform the statistical calculations desired by the users, the information center managers, and the suppliers of the data bases. Here will also be discussed any devices used to achieve maximum throughput speed or the servicing of the largest number of user requests in the shortest possible time. Software for data management operations will be discussed further in one of the following chapters.

Data Base Transformation routines are computer programs which are essentially outside of the information retrieval operation. These routines operate on data bases to transform them into a more desirable form or format or even order. Also included in this group of routines are any programs which extract certain information from a data base and form a new file with it (possibly to assist in the search operations). Data Base Transformation routines are not really necessary for the operation of a small information service center but become virtually essential when more than one data base is searched, since it is obviously desirable to have only one set of search programs. Software for data base transformations will be discussed in the next chapter.

The basic objectives to be considered in evaluating and selecting software for an information services and products system are (1) the minimization of total cost to the user, (2) the minimization of total response time or the time from the initial submission of the user's request to his reception of the desired information and (3) the provision to the user of a reasonably easy system to learn about and to operate. (31) It should be noted that these basic objectives are often contradictory or in conflict - - minimizing cost may make the system harder for the user to operate, etc. It is usually in attempting to achieve all these objectives that the problems associated with information retrieval software arise.

The problems associated with the software required to produce information services and products for outside customers appear to be divisible into two classes: problems associated with virtually any complex computer programming system that uses input from many different sources (magnetic tape, disc, teletypes) and that must scan a large body of data; and problems associated with producing software that is transferable. Problems that are associated with the former would be assuring the privacy and inviolability of each user's data and reorganizing or grouping the data and allocating computer memory and storage devices in an optimal way to the various data sources so as to secure the maximum throughput speed. These problems will be considered further in the chapter on data management software. As to the latter, when it is desired to transfer a group of computer routines to a data base, computer, or computer system different from that for which the software was originally written (as would be the situation for data bases that NASIC intended to use in-house), it is necessary to have software that is "transferable". For a certain collection of software routines to be transferable, they must possess characteristics such as being written in a common high-level language and modularity (or being composed of separable pieces). The characteristics required for software to be transferable will be described in detail in the last section of this chapter.

The criteria for selecting and evaluating software to be used by NASIC in providing the information services and products desired by their customers will now be presented. These criteria appear to be divisible into two main classes: general criteria or criteria that apply to any piece of software designed for any part of an information retrieval system and "transferability" criteria that apply when it is desired to transfer some piece of software to a data base, computer, or computer system different from that for which the software was originally designed. The second group of criteria would only be applicable when NASIC was considering using some data base in-house (or on a nearby computer on which NASIC had purchased time); whereas the first group of criteria would apply to any software employed by NASIC either directly or indirectly through an intermediary to provide information services for its customers. (These latter criteria include those used to evaluate software employed by an information center from which NASIC purchases some information services or products). The general criteria that apply to any piece of software designed for any part of an information service system are listed in Table 13 and are described more fully in the paragraphs below.

Documentation

Adequate documentation is obviously necessary for the software of any information retrieval system that NASIC might attempt to purchase or use in any manner. As a part of this documentation, all options available for each piece of software must be adequately described and illustrated. The documentation must include a complete description of all error conditions and messages, and a precise description of the steps necessary to correct each error condition. The latter is particularly important since NASIC personnel will have to identify and correct any error arising in the processing of a user's request. This documentation should also include extensive examples of each of the important types of request to assist NASIC personnel in the use of this system. (Documentation of an existing system is usually but not always available).

Reliability (Operability)

NASIC obviously does not want to use a system that is not very reliable in spite of any other considerations. But problems arise in determining just how reliable a system really

TABLE 13

EVALUATION AND SELECTION CRITERIA FOR INFORMATION RETRIEVAL SOFTWARE (GENERAL CRITERIA)

(Criteria ordered roughly in decreasing order of importance)

Documentation

Reliability (Operability)

Throughput time

Availability of updating capabilities

Cost

Availability of software maintenance from original supplier (Vendor guarantees of software reliability or future maintenance)

Degradation of response time with number of users (For time-sharing systems)

Operating time

Maximum number of users (For time-sharing systems)

Number of people wanting to use this particular software

Ease of usage

Uniqueness of software

is. It is certainly not adequate to simply accept the assurances of the software developer. Present or former users of this system should be asked their opinion of its reliability. Possibly NASIC should run an acceptance test on certain information systems to determine for itself their reliability. (It is assumed that any system under serious consideration by NASIC must be operable. NASIC would not at this stage want to consider any software that was still in the developmental stage. NASIC would certainly not want to have to debug or rewrite someone else's program).

Throughput Time

It is obviously desirable for NASIC to obtain that information retrieval system which has the minimum average throughput time (or minimum time from the submittal of the user's request to his reception of the output). But comparisons between the average throughput times for different systems are difficult. The throughput times for each of the four basic operations associated with information retrieval - current awareness services, batch retrospective searches, interactive retrospective searches and document delivery - are only comparable within these operations. Throughput times for the same data base for each of these four services will differ widely for the same system due to the widely differing nature of these four services. Throughput times, of course, also differ according to the data base used. Thus, when comparing throughput times between times for the same data base as well as the same basic operation. This implies that it will be hard to obtain a substantial body of data with which to compare system throughput times. Also, occasionally system A may have faster throughput times for current awareness searches but slower throughput times for batch retrospective searches than system B for the same data base, in which case the comparison between the throughput times for systems A and B would include a complicated tradeoff involving the comparative NASIC usage of current awareness services and batch retrospective searches. It is also possible that some installations may provide different throughput times for the same service on the same data base depending on the price charged the user. In these cases NASIC would have to analyze the trade-offs, though the final choice as to the option taken would probably be left to the ultimate customer. It should be pointed out that the throughput times compared should be those for the entire operation, not for various parts of the operation, since these would be largely meaningless.

Availability of Updating Capabilities

It is certainly necessary to have adequate software facilities available for updating or correcting any data base to be used by NASIC since most of the important data bases are being continually updated and corrected. But, because of this latter fact, it seems reasonable to assume that any group that has a data base will have some sort of facilities to update it. Thus, this consideration should not arise too frequently. As to the possibility of the software itself having the capacity of being updated (by adding new functions, operations or capabilities), this is probably not too important since it will often be possible in the future to find and use better, more flexible software on the common data bases. Also, it would seem that improving the capacity of the software would be something to be done in the distant future rather than now.

Cost

Costs will have to be determined for each software system and data base under consideration for each of the first three of the four basic uses or services: (1) Current awareness services; (2) Batch retrospective searches; and (3) Interactive retrospective searches. For each system and data base, it will be necessary to estimate the costs of doing a number of typical searches or operations for each of these three basic types of uses. It will be necessary to survey a number of suppliers of data base services to determine what sort of costs are typical and reasonable. However, the entire question of cost is largely inseparable from the questions of which data bases are being used and which capabilities are provided by the software. Thus, cost is less a software problem than an entire information services system problem, though there may be occasional instances in which basically equivalent software facilities operating on the same data base differ somewhat in cost, in which case cost will be the most important consideration.

Availability Of Software Maintenance From Original Supplier (Vendor Guarantees Of Software Reliability Or Future Maintenance)

To insure the reliability of software used by NASIC, or to correct any errors that might be found in such software in the future, it would be desirable to obtain some sort of assurances from the original supplier of any software used that it will be adequately maintained for the forseeable future. At this stage NASIC certainly does not want to take on the task of maintaining software, so any maintenance that must be done in the future should be guaranteed by the original supplier.

Degradation Of Response Time With Number Of Users (Response Time)

In interactive processing or processing under time-sharing, response time, the time between the submission of the user's request and the computer acting upon the request, is by far the most important and most basic component of both the total throughput time and the cost. Thus, when estimating either the total throughput time or the cost for a system employing time-sharing, the response time and its degradation with number of system users should certainly be considered first and emphasized most heavily in the analysis. However, response time will certainly differ widely for each of the two basic time-sharing operations - interactive current awareness service and interactive retrospective search, so that the comparison of response times between systems will certainly have to be done separately for each of these two services. Also, response time for each of these two services will differ widely according to the size of the data bases searched and the number of keys searched on.

Thus, it will be difficult to obtain a substantial amount of comparable data with which to compare the response times of different information centers. Also if the response time for interactive current awareness services for system A is better than that for system B, but the response time for interactive retrospective searches is worse, complex trade-offs will be involved in this decision.

Operating Time

This criterion does not refer to time-sharing or interactive processing, but refers simply to the amount of computer time required of a certain batch-processing operation. It is an important consideration not in itself but because it has an important effect on both the total throughput time and the cost for both current awareness services and batch retrospective searches. Thus, operating time will usually not be considered by itself, but will be considered in its effects on the throughput time and the cost for each operation.

Maximum Number Of Users (For Time-Sharing Systems)

This would seem to be of importance only in the distant future when NASIC might wish to service numerous requests for information from the same data base at the same time.

Number Of People Wanting To Use This Particular Software

This would have to be determined by taking surveys of possible users which is not the function of any software evaluation and selection criteria.

Ease Of Usage

Ease of usage is one of the less important criteria since NASIC personnel will often be the principal users of any data base and software system acquired by NASIC. Ease of usage will certainly have to give way in situations where a popular or useful data base can only be reached or can best be reached (best in the sense of providing the most relevant information at the lowest cost) through a hard-to-use software system. (The Ohio State University system might be rather hard to use since it requires the estimation of probabilities of relevance to the user). (33)

Uniqueness Of Software

This consideration should not arise too frequently in practice due to the intense competition in the software industry. It is rather unlikely that only one piece of software will be available to do any important particular job or function.

Questions relating to the fairness of the franchise arrangements or the number of present users of a certain system and who uses which parts of the system for how long or variations in the software usage cost with the amount of usage (the number of input requests) appear to be basically data base questions or to be intertwined with or intimately related to data base questions. A franchise arrangement would probably include the data base for which the software was written. Even if the franchise arrangement did not include the data base, the software would probably be specific to or intended for use on a certain data base in a certain format, and thus intimately involved with that data base. The number of present users, who they were, which parts of the system they used, and how long they used the system is obviously very highly dependent on the data bases employed and their popularity and usefulness. The usage cost of the software is obviously highly dependent on the size and format of the data base searched.

The criteria that apply to any software that NASIC wishes to transfer to a different computer, a different computer system, or a different data base are now presented in Table 14 and described more fully in the paragraphs below. It should be especially noted that the criteria of Table 14 are not a replacement for the criteria presented in Table 13 but are an addition to it for those situations and only those situations in which NASIC wishes to transfer software to another computer, another computer system, or another data base.

Dependence On The Particular Computer (Operability On A Widely-Used Computer, Machine Limitations)

In discussing the transferability of a software system, it is necessary to consider the degree of dependence of the software system on the peculiarities and limitations of the particular computer that it had been operating on - the characteristics of the particular central processing unit and the presence or absence of certain peripheral equipment, such as card or paper tape readers, card or paper tape punches, magnetic tape readers, disc storage units, printers, and their capacities, requirements, and other characteristics. To ensure the transferability of a software system the particular computer that it had originally been run on should be a commonly available computer with no unusual peripherals (for cards, tapes or disc storage) or unusual peripheral combinations or sizes. It is most important to have a sufficiently large amount of the correct type of storage. (The University of Pittsburgh system seems to be machine dependent. It will only operate on a PDP-10 with similar peripherals.)(32)

TABLE 14

EVALUATION AND SELECTION CRITERIA FOR TRANSFERABLE

INFORMATION RETRIEVAL SOFTWARE

(Criteria ordered roughly in decreasing order of importance)

Dependence on the particular computer (Operability on a widely-used computer, machine limitations)

Dependence on the programming language (Use of a universal or popular language)

Dependence on the particular operating system (Operating system availability)

Operability on different data bases

Modularity

Dependence On The Programming Language (Use Of A Universal Or Popular Language)

To obtain transferable software it is also necessary to consider the degree of dependence of the software system on the peculiarities and limitations of the particular language that the software is written in. To ensure transferability the language that the software was written in should be a common widely-understood high-level language like COBOL or PL-1.

Dependence On The Particular Operating System (Operating System Availability)

When discussing the transferability of a software system, it is necessary to consider the degree of dependence of the software system on the peculiarities and limitations of the particular operating system and compiling system that the software had originally been run under or compiled on. To ensure the transferability of a particular software system, the operating system that the programs have been run under and compiled on should be a common system without any peculiar adaptations or features. It is of course rather hard to define "common" for an operating system (or a computer or programming language) but it should be fairly obvious when an operating system (or a computer or programming language) is definitely peculiar or distinctly rare.

Operability On Different Data Bases

This is an important consideration since if some software is to be transferred it is certainly desirable to have the transferred software do as much for NASIC as possible, for it to operate on as many of the new data bases desired by NASIC as possible. (The phrase "new data bases desired by NASIC" could of course refer to data bases already used by NASIC but with whose software operating characteristics NASIC is unhappy, as well as data bases that NASIC wished to add to its collection). However, the direct applicability of software to data bases that were obtained from a source different from the software is improbable. Rewriting or modification of the program or conversion of the data base might be necessary, neither of which tasks NASIC might want to attempt itself.

Modularity

Modularity in construction refers to the division of a computer program into a number of small logically separate pieces. It appears that this is not too important a consideration for NASIC since it is doubtful that NASIC would attempt to rewrite or modify or correct any software it obtained during the next few years, which is where modularity is most important.

H. SOFTWARE CONSIDERATIONS: DATA BASE FORMATS

The basic reason for considering the conversion of a data base into another format, or even order, is that any information services center will almost always be operating on more than one data base. Since data bases originate from a large number of different sources and since no common form or format has become accepted by most data base producers, the data bases obtained by an information services center are liable to be in different forms and formats. But, since it would be a tremendous duplication of programming effort to write different search programs to search data bases in different formats, or even to write different search routines for inclusion in the same computer program to handle different formats, it is usually desirable for an information services center to convert all of its data bases into the same form and format. Because of the desirability of this conversion, it is necessary to consider the different formats in which data bases are written. It is also necessary to consider the problems raised in attempting to search each format, so that an information services center can choose the data base format which appears most cost-effective for searching.

A number of different standard record formats are used for various data bases. One of them is the PANDEX record format used for all data bases operated by the Mechanized Information Center (MIC) at Ohio State University in Columbus, Ohio (33).

Figure 4 shows the layout of a PANDEX record. It should be noted that the PANDEX record format makes no provision for a record directory, a group of words located at the beginning of a record indicating the locations of the start of each field of information in the record. It is fairly reasonable not to have a directory as long as each field of information is the same length from record to record. However, when fields are of varying lengths from record to record, as would probably be desirable if the information in the data base varied considerably in length from record to record, such a directory is very desirable. Otherwise it is necessary either to fill up the short fields in a record with filler to make them conform to the standard length for that field, or to search through each field of a record to find information located in the last fields. The former option is undesirable since it greatly increases the physical size of the data base. The latter option is of course highly undesirable since its use would increase the time required for searching tremendously. It would also make modification or correction of a record very hard and time consuming. Thus, in many cases directories in each record of a data base are very desirable and thus commonly used.

The MARC II data base containing bibliographic citations prepared by the Library of Congress is an example of a data base whose records contain a directory field. (34) The MARC II record format is shown in Figure 5. The record directory is composed of a series of fixed-length entries each containing the identification tag, the length, and the starting charac-

LAYOUT OF THE PANDEX (PX) RECORD

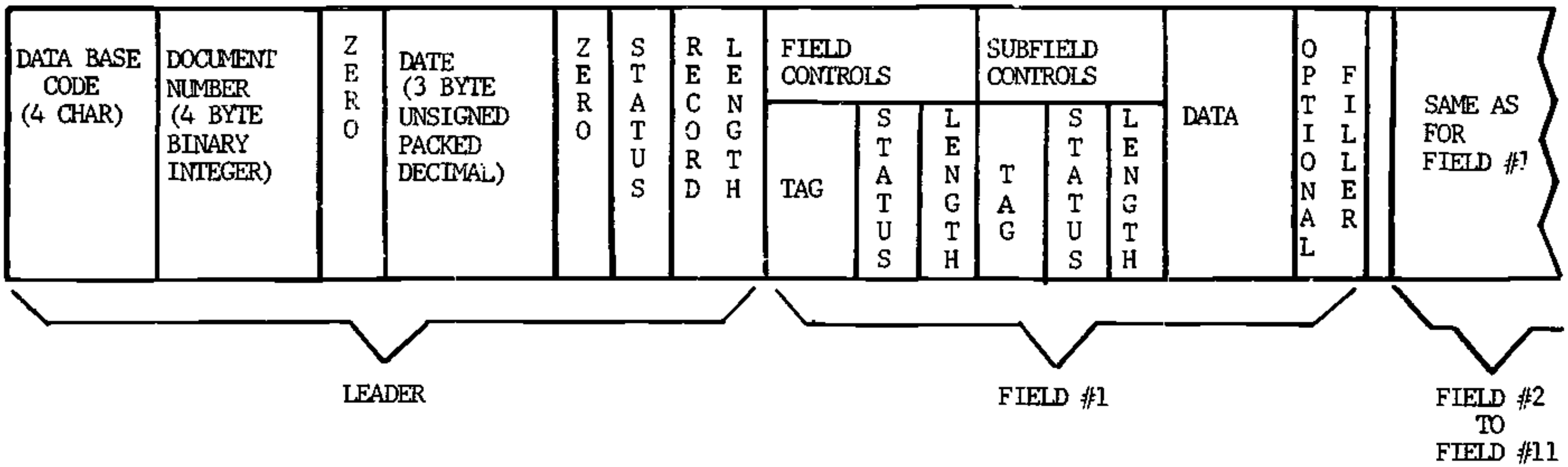

FIGURE 4

FORMAT OF THE MARC II RECORD

(A Record Containing A Record Directory)

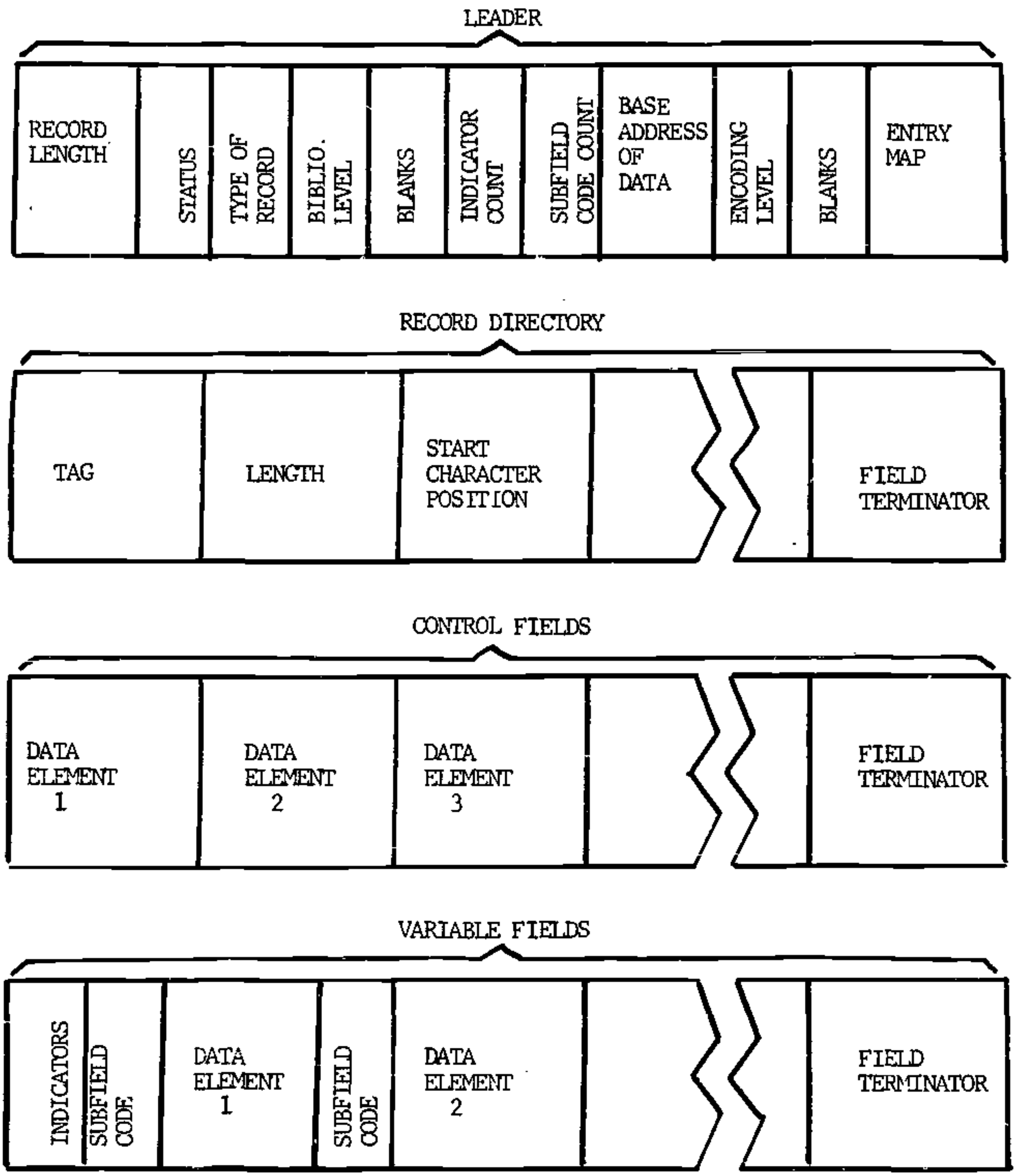

FIGURE 5

ter position in the record of each of the information (variable) fields.

Converting all the data bases in a center's collection to a common form or format to avoid duplication of the search routines is usually a fairly expensive and time-consuming effort. This is true not only because most of the data bases to be converted are fairly large but also because of the prevalence of errors in the data bases themselves or in the documentation describing these data bases. Thus, software to perform data base conversions will usually have to include extensive error checking features for the data bases. Even so, errors in the documentation for a data base or inconsistencies in its format and data representation will probably necessitate rewriting of the conversion routines and a number of false starts. The University of Georgia Computer Center, whose experience is probably typical, reports that for most of the data bases that they attempted to convert to their common format (The Standard File Format (SFF) of the Chemical Abstracts Services) the conversions had to be done at least twice, due principally to inadequate, inaccurate, or absent documentation or inconsistencies in the format and data representation of the data bases converted.(35) Inconsistencies in format and data representation are usually not detected until during or after the conversion, necessitating some modification of the conversion routines and then a second conversion attempt. Separate software will of course probably have to be written for each data base to be converted and checked for errors, so that writing the software for conversion of a large number of data bases is liable to be very time-consuming and expensive. Thus, the effort and expense of writing the appropriate software plus the difficulties involved in having to deal with inadequate or inaccurate documentation for the data bases and inconsistencies in their format and data representation imply that the conversion of even a small number of data bases to a common format will be very expensive and time-consuming.

Another reason for wanting to convert data bases to another form or format arises from problems caused by the hierarchy of concepts according to which a data base is logically organized. Logically (though not in actual practice) a data base is organized by a hierarchy in which broader, more inclusive index terms represent and include groups of more specific index terms. An example of such a hierarchy used to logically organize a data base is the following taken from the ERIC (Educational Resources Information Center) Thesaurus: "guidance personnel" is the broader or more inclusive index term which covers "adjustment counselors", "elementary school counselors", and "special counselors" as narrower index terms. (35) _If_ a data base were organized or indexed according to this hierarchy, "guidance personnel" would be the principal entry point or principal heading in a search and "adjustment counselors", "elementary school counselors", etc. would be the secondary entry points or headings under this principal entry point. If a search were done on a data base indexed in this manner, it would proceed by first locating "guidance personnel", the principal entry point, and then locating "adjustment counselors", say under this principal heading. But, since this method of moving down the hierarchy in a search would obviously take more computer time than if the area of interest could be located directly (especially if the hierarchy of organization is of more than two levels, unlike this simple example), in actual practice data bases are never ordered according to a hierarchy. Thus, it is usually desirable to have a data base indexed by the lowest level terms in the hierarchy, so that, in the above example, "adjustment counselors", "elementary school counselors", etc., would be the principal entry points. When the form, format, or even order of a data base is changed so that the terms at the lowest level of the logical hierarchy are the principal entry points, the data base is said to be inverted. The use of an inverted data base substantially speeds up the searching process, since the specific topics of interest can be located directly. (It is, of course, assumed that the user has some very specific area in mind and that this is clearly indicated).

Using an inverted data base to speed the searching process of course implies that the original data base and any additions to it are both put into inverted form before any searches are performed. But, inverting the original data base and additions to it require that same sort of data base conversion be performed on it. Such data base conversions to invert a data base are of course, subject to all the difficulties mentioned before in writing elaborate software to do the conversions and in performing the actual conversions; but, in addition, these data base inversions can involve a sort into some new order. Thus, inverting a data base will indeed make all subsequent searches on it less costly and less time-consuming, but inverting the data base (and additions to it) will usually be an expensive and time-consuming process.

I. SOFTWARE CONSIDERATIONS: SEARCH OPERATIONS

The search operations included in an information products and services system are of two major types: current awareness searches and retrospective searches. Current awareness searches consist of the matching of information in user profiles against the data in the latest issue of a data base and retrieving all the relevant information therefrom. Retrospective operations consist of a search through past issues of a data base for citations or material relevant to a given user request. Because of the differing sizes of the data bases involved and the differing states of the user's profile or request, current awareness searches and retrospective searches often differ quite widely in the manner in which they are performed and the problems encountered in conducting them.

Current awareness operations consist of searching the latest issue of some data base for all citations and material relevant to the specific areas of interest indicated on an already-available user profile. Since only the latest issue of a data base is searched in current awareness operations, and since updates for most common data bases are issued fairly frequently (every two weeks or every month), the size of the data base searched is usually fairly small. The profiles of the users are also already available and have usually been used in previous current awareness searches, so that no modification of them is contemplated or desired. The user profiles employed in current awareness searches are considered to represent precisely what the user wants. Because no modification of the profiles is permitted and the data bases are small, batch processing rather than interactive usage is usually quite acceptable for current awareness services. Failure to obtain the citations or information the user desires are uncommon due to the fact that most profiles have been used before in current awareness searches. Since current awareness searches are done when the data base is updated and since most data bases are updated fairly often, the user will not usually have to wait a long time for his results. The user's need for the information provided by a current awareness service is also usually less urgent than in other situations, since current awareness searches are used by researches principally to keep themselves up-to-date.

It is however virtually essential for the information service center to provide a fairly rapid response with current awareness services if only because the next current awareness search will be due shortly. An information center must put current awareness operations on a production basis if they are to be successful. Provision must also be made for the ultimate user to change his profile on the basis of previous current awareness search results given to him. Facilities for interaction with the user must be available and of adequate size to respond quickly to his new needs.

As to typical costs involved, the University of Georgia charges $5 to $10 per profile per current awareness search for a single data base. (32) Ohio State University charges $300 per profile per year for current awareness services which appears to be somewhat higher than the prices offered by the University of Georgia, since current awareness searches at the University of Georgia are often performed weekly or semi-monthly.

The current awareness operations of the University of Georgia will now be described. (32) The University of Georgia Information Center Staff perform the construction and coding of search profiles after interviewing the customer. The search profiles are then entered on an interactive system using CRT terminals. Syntax editing of the profiles is done immediately with diagnostics returned to the operator of the terminal. Profile updates can be performed as requested by the customer. The output of a current awareness search can be on either standard size paper or cards and will contain the title, primary bibliographic citation and any included index terms or codes. Current awareness searches are performed weekly, bi-weekly, semimonthly, monthly, or quarterly depending on the frequency with which the particular data base is updated.

Retrospective operations consist of a search through a part of the past issues of some data base, the number of years of issues to be searched being specified by the user, for all citations and material relevant to the specific areas of interest indicated by the user. Unlike the situation for current awareness searches, an interactive on-line system or semi-interactive system seems to offer substantial advantages to the user. When a user does a retrospective search, he does not have a previously-prepared and tested profile or request (unlike the situation that usually exists for continuing current awareness searches). The user is almost always making a retrospective search on a specific area of interest for the first time. Thus, it is very helpful to him if he can alter his request after trying it out on a small part of the past issues of the data base he wishes to search. By being able to modify his search request before he has done the entire search on the usually very large data base, the user saves both time and money since he does not have to pay for or wait for the results of a search that are of very little value to him, that do not contain precisely what he wants. Beyond this, it would certainly be very wasteful indeed if his search request were rejected because of a format or spelling error. Thus, because the user profile or request is usually not well set for a retrospective search and because the data bases to be searched retrospectively are very large, it is usually desirable for the user to be able to alter or negotiate his profile before he has searched all the desired issues of the data base. Retrospective searches can be totally interactive and on-line or they can be "semi" interactive on-line, as they are at the University of Georgia where correction and updating of the user's profile or request can be done interactively on-line. (The actual retrospective searching is however done off-line, but, the interactive on-line negotiation of the profile is certainly valuable for retrospective searching).

Since the data bases searched in retrospective operations are usually very large and since the searching is often done interactively on-line, it is necessary that the search operations themselves be as fast and efficient as possible. It is in searching large data bases that the advantages of file inversion, mentioned in the previous chapter as a method of speeding up the search operation, become most apparent. However, to obtain these advantages in speed, it is necessary to invert each addition to the data bases. Since each of the common data bases is updated fairly frequently, this means that much time and money must be spent to obtain inverted data bases.

As to typical costs involved, SDC, Systems Development Corp., charges \$43 per connect hour for on-line retrospective searches on the Chemical Abstracts Condensates (CA-C) data base and \$32 - \$35 per connect hour for on-line retrospective searches on their other data bases. For their semi-interactive retrospective searches (having a response time of about two weeks), the University of Georgia charges \$70 per profile per volume on many of their data bases such as Chemical Abstracts Condensates (CA-C) odd and even and Biological Abstracts (BA) ; but \$35 per profile per volume on other data bases such as Bioresearch Index (BIORI). However, due to the differences in the services provided, it is not reasonable to compare directly SDC's charges for retrospective searches with those of the University of Georgia.

SDC provides interactive on-line retrospective searches for past issues up to seven years old for some data bases. The files searched contain from 16 to 29 different categories of bibliographic information (author, keyword phrases, title, etc.). Some of these categories are directly searchable,but all categories can be searched on subsets of the file. All the information in a citation can be requested by the user through "print" commands. Response time for this service is usually almost immediate, but like other time-sharing systems degrades with increasing numbers of other currently-active users.

J. SOFTWARE CONSIDERATIONS: INPUT/OUTPUT PROCESSES, DATA MANAGEMENT REQUIREMENTS

This chapter will discuss the input/output routines and then the data management routines required by a typical information services and products center. This discussion will thus complete the consideration of all the different types of routines that form an information retrieval system.

The input routines are those portions of the entire information retrieval system which both permit and assist the user in constructing his profile or request for information. Profiles can be constructed in natural English prose as in the Lehigh University LEADERMART system[36] or must only use a restricted English vocabulary and/or logical indicators. Input routines that process natural English must of course have some capabilities for syntactically analyzing English sentences and must also have capabilities for searching thesauri for synonyms and narrower terms in the hierarchy for the words employed in the user's request. Such processors will often have the capability for interacting on-line with the user to negotiate his profile, to suggest related areas of interest to the user, or to force him to define his area of interest more precisely. Input routines that process requests written in a restricted vocabulary have to provide facilities to correct format and vocabulary errors in user requests. Such processors also often provide facilities for looking up synonyms and terms related to those given in the user's request, though sometimes the user is provided with a printed thesaurus to assist him in writing his request.

Input routines may also provide for profile inversion. Here this reduces to thesaurus look-ups to find the broader terms and the related terms connected with the user's request terms. (This assumes that the data base searched is ordered serially and is not in inverted form). This inverted profile is then matched against the data base to be searched on. Profile inversion can often be used as a substitute for file inversion to speed up the searching process. The use of profile inversion often eliminates the very costly and time-consuming process of file inversion.

Because of the above considerations, input software for an information products and services center can often consist of a number of elaborate routines which are costly to program. Employing the input routines can also contribute heavily to the cost of performing the entire operation.

The output routines are those parts of the information retrieval system which put the user's output into the format and order indicated by him and then make provisions for routeing this output to him. Output routines are usually "table-driven" processing routines. The user indicates as part of his input the format and order in which he desires his output. These indications are reduced to tables, one for each user. The output routines are written to be sufficiently flexibile to use these tables to format and to order the output as the user desires. (The output routines have to include sorting routines to put the output into the order indicated by the user). Thus, table-driven output processing provides the user with the maximum flexibility in output format and also permits many different orderings of the output items. The user is given maximum control over the form and order of his output.

The data management routines handle the complex details of the billing and accounting procedures used by information services and products centers. Information service centers almost invariably provide access to a number of different data bases which usually have different usage rates and rate structures and which derive from different sources. Billing each customer for the proper amount and performing the necessary accounting and computation of royalties is made very difficult by having all those different usage rates and rate structures. Thus, it is necessary to provide fairly elaborate and thus rather costly computer routines to perform all these complex billing and accounting procedures for each of a large number of users.

Included with this discussion of the data management routines are some comments on the provision of adequate privacy to each user and the system measures used to ensure maximum throughput speed-processing the largest possible number of user requests in the shortest possible time. The user wants his profiles and output information to be private and inviolate. He does not want any possible competitor viewing them and he certainly does not want anyone accidently or purposely destroying them. The problems of privacy and inviolability are two of the most serious problems faced by time-sharing systems including those operated by information service centers. In fact, the problems may be more severe for them due to the large number of users they often have and the fact that they are often connected up with terminals at remote locations.

The usual method of handling these privacy problems is assigning secret passwords or codes to each user and then including checks in the programs to ensure that only users with the right codes have access to the information corresponding to those codes. The University Of Georgia system assigns a secret code to each terminal and then attaches this code to all information coming from or going to that terminal, so that user access is only possible to information produced at the terminal he is using. (35)

One of the more obvious methods for an information products and services center to achieve maximum throughput speed is by grouping the searches for each data base together and then running each group separately only on the appropriate data base. However, this is only really successful when the center has a very large number of searches to perform in a short, but not too short, period of time. The center must have a fairly large clientele and must be able to defer some of its searches for a few days anyway, as would usually be possible with current awareness searches, for instance. Thus, this method of achieving maximum throughput speed is useful for some centers, but often inappropriate for others.

Another fairly obvious method of assuring maximum throughput speed is by the optimal allocation of computer memory and storage devices to the various sources of input data. Using disc storage for the data base information is probably the most efficient storage allocation for interactive on-line retrospective searches since large numbers of items must be readily available during such searches. For batch searching it is probably reasonable to have the data bases to be searched on magnetic tape, since the magnetic tape can be read in and searched serially.

K. COMMUNICATIONS ASPECTS OF INFORMATION SERVICE CENTER OPERATIONS

The successful operation of an information service center will be greatly dependent upon the nature of its supporting communications system. There will be the need by an information services center such as NASIC to communicate with an appreciable number of remote sources and data bases. At the same time there will be a growing number of queries addressed to an operational services center. Input queries may be modest in number at the outset but will eventually be very large.

A NASIC type center will use voice communication with its user customers and remote sources and service centers. It will need data transmission capabilities between terminals at the NASIC center and remote services centers. There will be slow-speed data transmission of query data and service center outputs. There will be medium and high-speed data transmission primarily of the quantity outputs from service center search operations. High-speed lines will be of particular interest when interactive processes requiring display terminals will be put into use.

NASIC will have the option to use direct dialing service or leased (dedicated) lines between communicating points. It can also make use of the services of commercial networks which has certain advantages and provides a way to consolidate certain costs.

A commercial network service usually includes costs of terminals and communications in the service fees. The discipline, procedures and protocols set up by a commercial network facilitate the process of accessing all other nodes on the network. This would be advantageous if a service center such as NASIC had a reasonable amount of traffic with these particular locations.

The NASIC operation will require the use of both dial-up and dedicated lines. A reasonable mix of these services can be established based upon the traffic needs that will evolve for the NASIC operation. Considerations which should be kept in view in determining when to use dial-up or dedicated line service are summarized in Figure 6.

Apart from communications with outside locations, a NASIC service center will have to deal with a difficult set of problems concerned with interconnections at its headquarters facility. A service operation will be at the center of input query activity, data base and data services accessing, as well as the output and delivery of search and processing results. At the beginning, staff personnel can probably manage to handle the interconnection problem, but as traffic increases this will not be adequate. There will be an insufficient number of staff people to do the job and the center's response capability will tend to deteriorate.

To deal with the interconnection problem at the headquarters facility, NASIC should consider the application of a "communications bus" approach. This is being developed and applied in different ways, but a good example of the kind of capability which can

CONSIDERATIONS IN DETERMINING USE OF DIRECT DIAL-UP OR DEDICATED LINE COMMUNICATIONS

DIRECT DIAL-UP

No line charges when there is no connection.

Faulty line problems are solved by re-routing

Data rates are low: 1800 baud (bits/sec) is a maximum for dial-up lines.

Quality of lines varies; a bad connection could mean many errors.

Cost reduction options are available (e.g., WATS service).

Response time to establish a connection is a factor.

Provides more accessibility for a service center to the user environment.

DEDICATED (LEASED) LINES

Use where message traffic is high, where remote locations are fixed, where average duration of calls is long.

Lines are available when needed, a physical connection exists.

Are high quality lines, tend to have lower error rates, cost more generally, can handle 50 kilobaud data rates.

To insure against failure a duplicate back-up line is required.

If line goes down, it usually takes longer to find a fault and correct it.

Is indicated when privacy is to be maintained.

FIGURE 6

achieved is given by the MITRIX system developed by the Mitre Corp.(38). MITRIX is a time-division multiple-access digital communication system employing cable technology with mini-computer controls and low cost components. It can be used in a closed loop configuration with the widest variety of communication devices, terminals, processors and data lines being capable of being connected to the central high-bandwidth "bus" channel. The opportunities to implement this system in modular fashion with modest initial cost, with abilities to automate interconnection functions as needed, make this an attractive approach to consider. A center such as NASIC will be very much concerned with keeping account of its service activity; it will also be billing service charges. There will be a wide variety of rate structures and in some cases royalty payments will have to be applied. For any operation with a sizeable volume, billing, bookkeeping and royalty controls can only be done with automated processing aids. With a system such as MITRIX in view for handling the internal headquarters interconnection requirements, it is possible to plan and implement the needed processing aids in a systematic way as the volume of business requires, all within reasonable economic bounds.

L. TECHNOLOGY CHANGE AND INFORMATION SERVICE CENTER PLANNING

As part of its initial plan, NASIC will be getting into operation with the use of "state-of-the-art" methods, equipment and procedures. There will be no opportunity in the early phases of NASIC to carry on extensive development effort or to do any extensive adaptation of newly available technology for NASIC requirements. It is the case, however, that NASIC is beginning its operations at a time when dramatic changes are taking place in the communication and processing fields. These developments are of great importance and they can have a strong impact on any NASIC system which will be set up. Although there will be limits on the actions which it may be possible for NASIC management to make at this time, they should be aware of key technology developments and they should take these into account in their forward planning. It will be useful to review briefly the status of technology in several areas which do have a direct relationship to a NASIC-type service center operation.

1. BRIEF OVERVIEW OF RELEVANT BUT CHANGING TECHNOLOGY

a. Speed, Power & Cost Trends In Computer Technology

The remarkable changes that are going on in the basic characteristics of computer tools is summarized in a study made by Ware several years ago(39). He listed and projected costs, speed, size and power factors for the period 1955 through 1975. The results of Ware's findings have been remarkably close approximations of what is actually happening. In a twenty-year period the size of the computer will decrease 10,000 fold for equal computational capability. In the same period the unit cost of computation is down [illegible] startling figure of 200,000 fold, while speed has [illegible] ed 40,000 fold. There has also been an explosive growth of installed capacity in the U.S. which, over the double decade 1955 - 75, will increase 160,000 fold. By 1975 computer processors should operate close to 10^9 operations per second. Ware's data is generally accepted as realistic, although the recent advances in the application of Large Scale Integrated Circuit (LSI) technology make some of his estimates very conservative. The clear implications are that there will be a continuing change in computer capabilities which will have a continuing effect on computer systems design, on usage patterns, and on costs.

b. Integrated Network Systems In the last few years there has been remarkable evolution of large integrated computer systems which usually consist of a network of processors, associated memories, and peripheral input/output equipment all interlinked(40). Often portions of these systems are separated by long distances. They may have huge data bases which can be accessed by hundreds to thousands of system users at locations which may be world-wide. The pressures to set up integrated networks are largely concerned with a need to support many system users and to keep costs of computer facilities within economic bounds. Integrated networks provide a way to share computer resources and to distribute processing capabilities as well as programs and data so that large groups of users can be served by a common set of capabilities.

It should be noted that the developing computer network systems are definitely tending to become utility systems. There is an emphasis to provide a spectrum of programs and processing functions which can serve for the common needs of most of the system users together with file storage capacity to serve many needs. Both batch and interactive processes can be supported in a single system. In order to achieve satisfactory response times and efficient time-sharing, the architecture of network systems tends toward distributed processing configurations[41]. That is, the central processor for the system is usually powerful and supports the large files required by the system, but many of the processor functions are carried out at the periphery of the system by smaller satellite processors and only occasionally is there need for these smaller processors to utilize the central portion of the system for new instructions, programs, or data. The processing load for a system of this kind is distributed and makes possible a reasonably economic configuration.

c. Fast, Low-Cost Communications Advances in the utilization of repeater relays in hard-wire communication lines, the refinement of digital communication methods including pulse-code modulation (PCM) has contributed to a major increase in digital communications capabilities and a rapid lowering of costs.

The full impact of this will be realizable in the next few years. For example, a communication channel of 1.5 megabit capacity for a 25 mile data-link, a short time ago, could be obtained for one-fifth of the level of what the cost of a 50K bit channel had been for some time previously. It has been estimated that the eventual cost for a data-link from a terminal to any location in the U.S. will level off at about $.30 per hour of usage(42).

The low level of communication costs for a computer network system has major implications for the design of systems. Whether a computer is located in the next room or one hundred miles away will make relatively little difference in the overall system cost(43).

d. Large High-Speed Memories A new set of processing capabilities which will be available soon and which can have a major impact on system design are represented by high-speed mass memories(44).

There are memories in experimental operation with capacities of the order of a trillion bits(10^{12} bits). As a reference it may be noted that the complete textual holdings of the Library of Congress can be represented by 10^{14} bits. There are various media used for storage, different recording and reading methods are used, and several companies will be marketing particular mass memory products in the next year or two. It is clear the good reliability can be achieved, the cost of the memory unit is at a reasonable level, and good average access times are being achieved (of the order of present disc memory access times). There are several specialized mass memories which are used in combination with powerful central processors where transfer rates of the order of 40 megacycles (40 x 10^6 bits/second) are attained.

The availability of mass memories of high capacity, good access times, good reliability at reasonable cost will very much enhance the response times which are possible in distributed network systems because of the abilities to transfer large streams of programs, instructions, and data at high speed. With mass memories available, problems concerned with input processing, labeling, and structuring of data to be stored will be intensified. It is quite apparent at this time that we do not know how to store data in a mass memory so that we can use the storage efficiently, even if we could afford the cost of input and take all the time needed to load the store. In any case, it seems certain that the impact of new mass memories on system design will be a major one.

e. Low-Cost Computer Processors In line with the general trends mentioned above, there has been a large output of small, powerful low-cost computer processors in the past few years. These machines have been generally labeled "minicomputers" and are capable of performing a wide spectrum of processing functions. They can be used as stand-alone processors or as satellite processors in a network system and can be switched from one mode to the other. Beyond the minicomputer in processor size is the "microcomputer". These computers are a development of greatest importance in the computer field and will have a profound impact on machine processing systems for many years to come(45). Microprocessors or microcomputers are an application of Large Scale Integrated Circuit (LSI) technology which is usually implemented via semiconductor technology. In very general terms LSI consists of the use of hundreds or thousands of circuits or memory cells in a single integrated unit on a minute chip. Note that memory as well as computing capability can be implemented in LSI.

LSI developments have been watched with much interest over the past few years but the picture that is now evolving and which brings new implications is that much higher levels of computing power will be achieved in LSI and these will be implemented at cost levels which were not considered possible in previous estimates. Not only will low-cost computing power be available, but large amounts of high-speed memory will also be available in LSI at unbelievably low costs. Processors with very high computing power will be producible as separate LSI units; high-speed memory units in LSI with millions of bits of capacity will also be available. All of these LSI units will be capable of being produced in large quantity lots at costs of, say, less than $100 per processor or memory unit.

f. Low-Cost Display Terminals A display terminal has advantages over the terminal with only a keyboard because it makes possible a much greater "bandwidth" between a human user and the machine system. A computer can feed back much more data faster, with the use of a display, and this facilitates certain interactive man-machine processes. Costs of display terminals are continuing to become more reasonable. Where costs of a display terminal were at the $25,000-$50,000 level a short time ago, [illegible]ral are available at about $2000 and will probably level off eventually at about the cost of a typewriter. The availability of low-cost display terminals has a major relationship to the design and utilization of the machine processing system.

It should be noted that many particular processing functions can be incorporated in a display terminal so that it becomes a kind of small specialized processor. These terminals have been designated as "intelligent terminals" and are significant because they provide a way to relieve a central support system of a processing load which might be required because of a specialized application.

2. THE POTENTIAL OF TECHNOLOGY ADVANCES FOR NASIC SEARCH CENTER OPERATION

There is no disagreement that many interesting technological developments are coming along and that these will at some time be put to useful applications. However, there is a difference of opinion as to how much a service center, such as NASIC, which is just beginning operation, should pay attention to evolving technology. The argument could be made that NASIC should get under way using established procedures and tools in order to get a beginning level of service started even if this is largely a minimal activity. There is another viewpoint which says that much of the newly developing technology is already at a stage where it can be applied and adapted and that what is needed is a clear specification of system needs and requirements in order to get on with the adaptation of the new technology. This viewpoint also holds that costs of application and adaptation will be within reasonable economic bounds if the benefits and the increased system outputs and services are taken into account. There is not much to be gained by further discussion of these different viewpoints but there is some value in pointing out that NASIC would tend to gain appreciably in the quality and capacity of its operation if some of the projected technological goals are fully realized.

In the future, improved communications would include lower costs, high-speed wide band-width channels, automated switching, networking and interconnection processes. Improvements in data storage capabilities would include lower cost memories, memories of massive capacity, high data transfer rates, automated file organization. Improved on-line processing would include low-cost display terminals, stand-alone display processors, a wide-spread proliferation of terminals, improved integrated operating systems, distributed processing networks, low-cost, small size mini and micro processors. What do these developments mean for a NASIC service center operation? What potential do they represent?

The availability of greater capacity for making storage at lower costs means an ability to replicate holdings, to store massive libraries of data and information at a local center. This means a new set of options to provide faster processing and service. Low cost microprocessors provide new opportunities to do multiple searches and processing operations to accommodate larger volumes of queries for more system uses. The facilitation of on-line capabilities provides new levels of accessibility. Not only will more people with information needs have the ability to explore the holdings and services of the center, but also there will be new ways to provide delivery of search results and document texts. It all adds up to new potentials to provide higher levels of accessibility and an improved ability to handle higher volumes with fast response at costs which will be acceptable to the user population.

3. IMPLICATION FOR NASIC PLANNING POLICY

The preceding discussion regarding technology change and its potential for NASIC has been intended to underline the fact that the essential tools which NASIC needs for its operations are in fact undergoing significant change. In these circumstances it is important to emphasize that NASIC should not use its resources to establish and elaborate procedures and processes which are certain to be altered or eliminated in a fairly short time period. Where new approaches and tools with a prospect for stability are proven to be satisfactory and reliable they should be used and exploited.

It may be advantageous for NASIC to set up certain experimental subsystems in order to provide a reasonable interplay with new and evolving capabilities. For instance, it would be possible to select some segment of the user population which has a clearly defined set of data and information needs. Experimental services could be provided using all possible tools including newer methods to give the most rapid response, the best coverage, and the fastest document delivery. It might be important to set up direct communication lines with selected users to test certain on-line interactive interrogation and searching of selected files.

Consideration should be given to the use of the NASIC operation as a test-bed situation where new developments could be tried and tested. An active part of the NASIC program should include a certain amount of on-going research and development pointed toward the achievement of NASIC's basic goals. There is no reason why R&D effort could not reinforce and supplement the primary objective of NASIC to serve the Northeast Science Community using existing methods and sources.

In setting up a new operation it would be an unwise approach to concentrate wholly on an operation which would employ only "state-of-the-art" methods and means. It is doubtful that any operation based on this approach would be able to meet requirements for coverage, speed of response and quality of service especially if a sizeable user population is to be served. NASIC will want to anticipate that its operation, after a period of initiation and growth, will be handling a large volume of queries and responses. It will be serving a large user world and will have to move in the direction of automation of its handling, search, processing and delivery processes in order to give a continuing level of satisfying service.

M. SOME MANAGEMENT ASPECTS OF AN INFORMATION SERVICE CENTER OPERATION

An information service center as proposed in the NASIC program is intended to serve as a kind of brokerage and interface between individuals in the science community who have information needs, and the sources, data bases and services which can serve these needs. By the aggregation of data bases and services and the providing of access to services at one central point, it is intended that certain advantages be realized for the center operation as well as the potential users.

If the objectives for NASIC are to be attained, it will be essential that NASIC management be very clear about the kind of user market it intends to serve, as well as the nature of the service it intends to provide. In the following discussion we believe it is useful to review what options NASIC may wish to exercise.

Objectives for Continuing Support The NASIC project is in a beginning phase. The general guidelines which have been set up emphasize the desirability of achieving a non-subsidized, paying operation within a three year period. A major objective, then, is to provide services which will attract customers who are able to pay the costs of the services given. This is of prime importance because, without the prospect of subsidization after the three-year period, the very survival of a NASIC operation is at stake. In its decisions about kind of service and the choice of user clientele, the need then to generate a paying operation will be a major factor.

User Market. It is essential that NASIC make an early decision about the user clientele it intends to serve. The charter for the NASIC project does make clear that NASIC has a prime responsibility to the academic science community. Students and faculty members of the many colleges and universities of the Northeast will be interested in NASIC services. There is a question, however, whether many individuals in the academic world will be willing or able to pay for the services which are offered. Institutional contributions will undoubtedly be an important source of NASIC funding when NASIC has demonstrated its usefulness. However, it is very doubtful that this support alone can provide what NASIC needs. It is almost inescapable that NASIC will have to point to the commercial and industrial elements of the science community as a primary marketing objective.

Kind of Service. An information services center will provide a spectrum of data and information services as described in earlier sections of this report. To achieve a viable operation which will be attractive to potential users in the science community, it is important that NASIC adopt a clear policy regarding the kind of service it intends to give. It is not sufficient to assemble all available data bases and information files together with a means to access them, and then expect customers to approach the service center to use what has been assembled. Instead, it is important that the service center management take steps to understand what the information need is of the typical user in his environment, and then take steps to focus the service center operation to meet that need. It will be profitable to spend time and effort to identify as clearly as possible what the information and data needs of individuals are and then adjust the objectives of the service center operation on the basis of this knowledge. There will be opportunities to be selective about the choice of data bases and sources. There will be options about kind of access and document delivery. What is to be provided by the center should be the best service available to meet in the best and fastest way a real and unsatisfied customer need.

Promotion. It is a reasonable extension of the preceding discussion to emphasize that NASIC management will wish to give attentive consideration to all kinds of means to promote NASIC services. It is the case that there are many situations in both the academic and commercial areas of the science community where data and information requirements exist and are not being served. NASIC will want to publicize vigorously what services and data sources it can make available and the convenient ways to access these services. A strong promotional program will be a necessity to assist in making a NASIC center a viable operation.

REFERENCES

1. Lancaster, F.W. - "The Cost-Effectiveness Analysis of Information Retrieval and Dissemination Systems", Journal of the American Society for Information Science, vol. 22, no. 1, Jan.-Feb., 1971, pp. 12-27.

2. Marcus, R.S. et al - "The User Interface for the Intrex Retrieval System", In Walker, D.E., ed. - Interactive Bibliographic Search: The User/Computer Interface. Montvale, N.J., AFIPS Press, 1971, pp. 159-201

3. Soracevic, T. - "Comparative Effects of Titles,Abstracts and Full Texts on Relevance Judgements", Proceedings of ASIS, vol. 6, 1969, pp. 293-9.

4. Lancaster, F.W. - "Evaluation of the MEDLARS Demand Search Service" Bethesda, Md., National Library of Medicine, 1968

5. Lancaster, F.W. - "Interaction Between Requesters and a Large Mechanized Retrieval System", Information Storage and Retrieval, vol. 4, no. 2, June 1968, pp. 239-252

6. Carroll, K.D. - Survey of Scientific - Technical Tape Services, New York, American Institute of Physics, 1970

7. Landau, H.B. - Research Study into the Effective Utilization of Machine-Readable Bibliographic Data Bases. Philadelphia, Auerbach Corporation, 1970

8. Finer, R. - A Guide to Selected Computer-Based Information Services. London, ASLIB, 1972

9. Schneider, J.H. et al - Survey of Commercially Available Computer-Readable Bibliographic Data Bases. Washington, D.C., ASIS, 1973

10. Patrinostro, F.S. & Mulheim, N. - Available Data Banks for Library and Information Services. Tempe, Arizona, the LARC Association, 1973

11. Survey of Bibliographic Retrieval Systems With Machine-Readable Data Bases. Silver Spring, Md., Computer Sciences Corporation, 1972

12. Davison, P.S. & Matthews, D.A.R. - "Assessment of Information Services", ASLIB Proceedings, vol. 11, July 1969, pp. 280-3

13. Bourne, C.P. - Characteristics of Coverage by the Bibliograpy of Agriculture of the Literature Relating to Agricultural Research and Development. Palo Alto, Information General Corporation, June 1969, PB 185425.

14. Martyn, J. - "Tests on Abstracts Journals: Coverage, Overlap and Indexing", Journal of Documentation, vol. 23, March 1970, pp. 45-75

15. Bradford, S.C. - Documentation. London, Crosby Lockwood, 1948.

16. Fairthorne, R.A. (RNE) - "Empirical Hyperbolic Distributions (Bradford - Zipf - Mandelbrot) for Bibliometric Description and Prediction", Journal of Documentation, vol. 25, no. 4, December 1969, pp. 319-43

17. Zipf, G.K. - Human Behavior and the Principle of Least Effort. Addison Wesley, 1949

18. Wood, J.L. et al - "Overlap in the Lists of Journals Monitored by BIOSIS, CAS, and EI", Journal of the American Society for Information Science, vol. 23, 1972, pp. 36-38

19. Flanagan, C.M. - "Coordination - a Detailed Review of the Relationships Among the Publications and Services of BIOSIS, CAS, and EI", Journal of Chemical Documentation, vol. 13, no. 2, 1973, pp. 57-59

20. Bourne, C. P. - Overlapping Coverage of Bibliography of Agriculture by 15 Other Secondary Services. Palo Alto, Information General Corporation, June 1969 PB 185069

21. Hansen, I.B. - "CA Condensates as a Retrospective Search Tool: a Commentary", Information Storage and Retrieval, vol. 9, no. 4, April, 1973, pp. 201-205.

22. O'Donohue, C.H. - "Comparison of Service Centers and Document Data Bases - a User's View", Journal of Chemical Documentation, vol. 13, no. 1, 1973, pp. 27-29

23. Devon, T.K. et al - "Comparative Evaluation of Ringdoc and CBAC", Journal of Chemical Documentation, vol. 13, no. 1, 1973, pp. 30-32

24. Montgomery, R. R. - "An Indexing Coverage Study of Toxicological Literature", Journal of Chemical Documentation, vol. 13, no. 1, 1973, pp. 41-44

25. Ashmole, R. F. et al - "Cost Effectiveness of Current Awareness Services in the Pharmaceutical Industry", Journal of the American Society of Information Science, vol. 24, no. 1, 1973, pp. 29-39

26. Sturdivant, C.A. - Geoscience Information Society Newsletter, December 1972

27. Williams, M. G. and A. K. Stewart - ASIDIC Survey of Information Center Services, Chicago, IIT Research Institute, June 1972

28. Park, M. K. et al, - "Education in the Use of Modern Information Retrieval Techniques", Journal of Chemical Documentation, Vol. II, No. 2, 1971, pp. 100-102

29. Beauchamp, R. O. Jr., et al "Comparative Searching of Computer Data Bases," Journal of Chemical Documentation, Vol. 13, No. 1, 1973, pp. 32-35.

30. Scott, E. J. et al - "A Technique for the Evaluation of a Commercial Information Service and Some Preliminary Results from the Drugdoc Service of the Excerpta Medica Foundation," Information Storage and Retrieval, Vol. 7, 1971, pp. 149-165

31. Aron, J.D., "Information Systems in Perspective", Computing Surveys, vol. 1, no. 4, December 1969, pp. 224-225.

32. Marron, B. Fong, E., Fife, D.W., and Rankin, K., A Study of Six University-Based Information Systems, U.S. Department of Commerce, National Bureau of Standards, Washington, June, 1973.

33. MIC Bibliographic Retrieval Software Technical Manual, The Ohio State University Mechanized Information Center, Columbus, Ohio, August, 1973.

34. Annual Report of the Mechanized Information Center, The Ohio State University Mechanized Information Center, Columbus, Ohio, March, 1972.

35. Final Report to the National Science Foundation on Grant GN-851 to Expand the University of Georgia Information Center, The University of Georgia Computer Center, Athens, Georgia, Dec. 1972.

36. Kasarda, A. J., and Hillman, D.J., "The LEADERMART System and Service", Lehigh University, Bethlehem, Pennsylvannia, 1972

37. Coltman, Robert - "Toward the Wired City: Telecommunication at Mitre", Mitre Matrix, 5, No. 6, 1972, pp. 39-52.

38. Mitre Corp. - Report M71-113 Digital Cable Communications, 17 pp. January 1972

39. Ware, W. "The Computer in Your Future" Rand Corporation, Report P3626, November 1967.

40. Schwartz, M., Boorstyn, R. R. and Pictsholtz, R. L., "Terminal-Oriented Computer-Communications Networks", Proceedings of IEEE-60, pp. 1408-1422, 1972

41. Chen, T. C., "Distributed Intelligence for User-Oriented Computing", AFIPS Conference Proceedings FJCC, pp. 1049 - 1056, 1972.

42. Estimate by Dr. H. McDonald, Bell Telephone Laboratories, "Conference on Multiple-Access Computer Networks", April 20-22, 1970.

43. Goodwill, D.Z., "A Look At The Future Impact Of Computer-Communications On Everyday Life", Technology Forecast and Social Change - 4, pp. 227-233, 1972.

44. Hoagland, A.S. "Mass Storage - Past, Present and Future", AFIPS Conference Proceedings, FJCC, pp. 985-991, 1972.

45. Dennis, S. F. and Smith, M. G., "LSI - Implications for Future Design and Architecture", AFIPS Conference Proceedings, SJCC, pp. 343-351, 1972.

46. Meadow, C. T., The Analysis of Information Systems, Melville, Los Angeles, 1973.

47. Salton, Gerard, Automatic Information Organization and Retrieval, McGraw-Hill, New York, 1968.

ATTACHMENT 1

THE PRESENT STATUS OF ON-LINE INTERACTIVE RETRIEVAL SYSTEMS

by F. W. Lancaster

Effectiveness And Cost-Effectiveness Considerations
For NASIC Information Services Operation
Q.E.I., Inc. Bedford, Mass. October 20, 1973

Introduction

This report is based largely on an important and unique workshop held by the Institute for Communication Research, Stanford University, April 23-25, 1973. The Institute has a contract with the National Science Foundation to conduct a comprehensive, comparative analysis of some major on-line, interactive systems for bibliographic searching in use in the United States at the present time. The workshop itself was considered an important element in this overall evaluation.

The principal investigators are Ed Parker and Tom Martin of the Institute for Communication Research. A committee of "experts" has been formed to provide assistance and guidance in the conduct of the study. The members of this committee, which includes the present writer, are listed in Appendix 1. Members of this committee acted as moderators of the five sessions of the workshop. Eleven systems are being studied. Brief "snapshots" of these systems are given in Appendix 2 and the system representatives attending the workshop are listed in Appendix 3, along with other observers who were invited to the sessions.

The eleven systems include the major bibliographic searching systems now in use with two notable exceptions: the New York Times Information Bank and the INQUIRE system. It was not possible to include the former because of the proprietary nature of the system, which was designed by IBM exclusively for use by the New York Times. The owners of INQUIRE chose not to participate in the study.

While the study is largely concentrated upon general-purpose bibliographic searching systems, some special-purpose systems were included for purposes of comparison. One special-purpose system is RIQS, the Remote Information Query System, developed at Northwestern University. RIQS is designed largely to allow individual academic researchers to set up, manipulate and query their own personal collections of documents, data, or bibliographic references. RIQS may be regarded as both a bibliographic retrieval system and a management information system. The NASIS system, operated by the NASA Lewis Research Center, is designed to be compatible with RECON and DIALOG. The major NASIS data base is a large collection of aerial photographs, but other data bases, including a bibliographic data base from the Nuclear Safety Information Center and some management data bases, are also included in the overall NASIS system. All of the systems represented, with the exception of LEADER, are designed to respond to conventional Boolean searching strategies. LEADER operates in a completely different way. Document representations in LEADER consist of noun phrases, extracted from text by computer. The system may be queried by means of a request in the form of an English sentence. This sentence is matched against the stored document representations (in noun phrase form) and those representations that best match the request are displayed for the user (i.e., the search process is essentially a pattern matching operation). DATA CENTRAL, STAIRS, LEADER and INTREX were designed primarily to operate with natural language. ORBIT, DIALOG, BASIS-70 and RECON were designed primarily to operate with controlled vocabularies, but each has been modified, or is in the process of being modified, to allow searching in a text mode. The other systems operate with text or limited vocabulary, dependingupon specific applications. RIQS has no capability for constructing and operating on inverted files.

The system "snapshots" of Appendix 2 present basic features of these systems in a standard format. For each the following characteristics are given:

(1) the type of data base handled,

(2) whether or not the system has a text search capability,

(3) system availability (accessibility),

(4) type of user,

(5) type of terminal,

(6) type of service offered,

(7) whether the terminals are in a central location (i.e., in a center of service) or distributed remotely (personal terminals),

(8) type of charges applied,

(9) programming language used, and

(10) hardware for which the system is designed.

The workshop itself was unique in that it brought together representatives of eleven systems, including competitors, in order to exchange views and experiences. The major objective of the symposium was to try to arrive at a set of "minimal features" that all participants agreed were essential to the operation of an efficient, viable interactive system. The meeting was remarkably effective in that agreement was generally reached on most of the features discussed. All of the systems, with the exception of LEADER, were demonstrated at the meeting.

The meeting was divided into six sessions, each with its own moderator, devoted to the following topics:

(1) The searcher/task environment

(2) The data base environment

(3) Search negotiation features

(4) Display and secondary support features

(5) Instructional and diagnostic features

(6) Overview and future directions

The later sessions of the workshop brought forth detailed and constructive discussion and they are worth summarizing here for that reason. The summaries of these sessions presented the most complete and up-to-date data now available on the characteristics and capabilities of present-day on-line, interactive retrieval systems.

THE SEARCHER/TASK ENVIRONMENT

This was the subject of the first session of the symposium. The discussion was general and exploratory. Participants were feeling their way only. The moderator, Nance, pointed out that the published literature on the development and operation of on-line systems for information retrieval is very sparse. He also suggested that most of the eleven systems represented at the symposium were developed in relative isolation, without direct reference to other existing systems. If many systems, developed independently, incorporate very similar features it would suggest that there is some evidence of consensus on the desirability of or need for the features. An important function of the symposium was to determine how much agreement exists as to which features should be regarded as "minimal" and which should be regarded as desirable, although not "minimal."

The later sessions of the workshop, which got down to details and which form the real heart of this report, did explore the degree of consensus existing among the various participants.

THE DATA BASE ENVIRONMENT

This second session, chaired by Belzer, also dealt largely in generalities. It attempted to explore the question of how much the user needs to know about the data base he wishes to interrogate. Each representative described various characteristics of data bases handled by his system. This portion of the symposium presented necessary background for the one and one-half days of discussion that followed, but it is not worth reporting in detail here.

SEARCH NEGOTIATION FEATURES

The major search negotiation features are presented in Table 1, with various symbols supplied to indicate which systems have which features. The starred (*) features were regarded by the majority of participants to be "minimal features" (i.e., features without which an on-line, interactive system is unlikely to function effectively).

Request sets

This refers to the capability of assigning a number to each line of a search statement (e.g., a string of terms in a Boolean OR relationship) and allowing the building of complete and complex search strategies by incorporating logical sets by line number, as in the following example:

1. A or B or C or D or E
2. Q or R or S
3. 1 and 2
4. 3 and not M

There was general agreement that this capability should be regarded as minimal. Only the representative of LEADER, which operates in a non-Boolean search mode, disagreed.

Dictionary access

This refers to the capability of displaying, in alphabetical sections, the "vocabulary" of the system, whether the vocabulary be a controlled vocabulary (subject headings or a thesaurus), text words, or indexing phrases (as in LEADER or INTREX). As examples, the NEIGHBOR command in ORBIT and the EXPAND command in RECON will cause the terms alphabetically adjacent to the input terms to be displayed. The alphabetical display is generated even if the input term does not appear in the data base (i.e., the terms closest to it alphabetically are displayed); this allows the system to compensate, to a certain extent at least, for some spelling or typing errors. This is perhaps particularly useful for names of authors. In DATA CENTRAL and SPIRES the alphabetical display does not show the number of postings associated with each term in the display; in the other systems having the dictionary access capability (see Table 1) tallies (postings) are displayed.

Related term capability

This refers to the ability to request, for any search term entered, that the systems display terms that are semantically related (e.g., see also or RT relationships). Clearly, this feature implies a system that makes use of a controlled vocabulary, although LEADER has its own form of related term capability (it will display phrases that most closely match a sentence input by the searcher). In DIALOG the feature is turned on automatically. That is, when a searcher enters a term the related terms are displayed automatically. In other systems the searcher requests the related terms display. DIALOG, RECON and LEADER number each term, so that they may be incorporated into the search strategy by number. STAIRS automatically adds each related term into the request set unless overridden by the searcher. This feature was regarded as very useful but not "minimal" by most participants.

Hierarchical thesaurus

This relates to the capability of being able to display a portion of a hierarchical thesaurus, where such a highly structured thesaurus exists in the system. For example, the command TREE in ORBIT will cause a display of the complete hierarchy in which a particular term appears. Again, regarded by most participants as desirable but not minimal.

Incorporation of synonym tables or term hierarchies

This is a related capability. It refers to the ability to incorporate easily into a search strategy a complete block of related terms. The best example is probably the use of EXPLODE in ORBIT, which causes the complete hierarchy beneath the "exploded" term to be incorporated into a search strategy. In DIALOG and RECON, the searcher can incorporate a whole block of displayed terms by using a "range" indicator (e.g., E 05 - E 17). In DATA CENTRAL stored synonym tables can be pulled into a strategy by unique identifying number. In STAIRS the searcher is given an automatic synonym expansion capability. That is, all synonyms stored by the system will automatically be pulled in when the user enters a search term, unless this feature has been overridden by the searcher.

Search field control

This was regarded as a minimal feature, especially for systems searching highly formatted records. It refers to the capability of specifying which field (e.g., title, abstract) in a record is to be searched. An X in the Search Field Control column of Table 1 means that the system will search all fields unless the user specifies that a particular field is to be searched. A ✓ means that the default condition assumes that some particular field (e.g., descriptors) is to be searched unless the user specifies otherwise.

Boolean operators

The ability to combine terms using the Boolean AND, OR and NOT operators is a minimal feature for all systems except LEADER, which functions in a completely different search mode. In ORBIT, DIALOG, RECON, BASIS-70, NASIS and RIQS the logical AND is processed before the logical OR. DATA CENTRAL processes AND before OR, but processes OR before &. In STAIRS and SPIRES the leftmost operator, whichever it is, is processed first. In INTREX no compound statements (both AND and OR in a single statement) are allowed. There was general agreement that some form of standardization of Boolean operators was needed, particularly to allow an on-line user to move freely from one system to another. For the transfer from one system to another in a completely mechanized way the problem does not exist because mapping algorithms can be developed to take care of system differences of this type.

Word proximity operators

This relates to the capability of being able to specify that a particular combination of terms, on which a search is conducted, should occur within a specified distance of each other in the document record. This distance may be specified by structural unit (same paragraph, same sentence) or by actual word distance (two words must occur immediately adjacent or no more than x words apart). The feature is clearly minimal for full text systems but not for systems operating on sets of index terms. It was generally agreed that, for most purposes, strict word distance was probably the best method to achieve this system capability. Very little need has been found to specify word co-occurrences within linguistic units (paragraph, sentence). All systems represented here allow exact phrase matching (i.e., two or more words must occur in adjacent positions in text). In ORBIT and SPIRES this exact phrase matching is achieved by sequential search. Only DIALOG, RECON and DATA CENTRAL allow the user to specify how many words may separate two search terms.

Arithmetic operators

This capability involves the use of GREATER THAN, LESS THAN, and BETWEEN operators. It is particularly valuable for systems handling numerical data (e.g., dollar values) and was regarded as a minimal feature by most participants. In STAIRS the arithmetic operators can only be used when performing a sequential search.

Suffix removal

There was general agreement that an interactive system, particularly one handling full text, should allow suffix removal (truncation). Usually suffix removal is under the control of the user. He enters any root and follows it with some truncation code. In INTREX, however, search terms are automatically "stemmed" unless the term is followed by an exclamation point. For example, the term "surgery" is automatically reduced to its root, "surg...", and will retrieve all words with this root, whereas "surgery!" retrieves only this particular character string. In DATA CENTRAL a final "s" on a word is automatically disregarded.

It is also desirable in many situations to be able to search on strings of characters other than prefixes (i.e., suffixes or infixes). Searching on suffixes can be a very powerful device in a natural language system. This type of search presents no particular problem in a batch processing, serial searching operation. It does, however, present problems in an on-line, random access system based on inverted files. It would certainly be extremely expensive to maintain inverted files for all suffixes, as well as prefixes. However, there is no real reason why inverted files could not be set up for selected suffixes that are known to be especially valuable in searching particular collections (e.g., ...MYCIN).

Phrase decomposition

This is an unusual feature possessed by only three of the eleven systems. It involves the decomposition of a natural language phrase into "significant" words, ignoring nonsubstantive words. In INTREX a logical AND relationship is assumed between the remaining words. In SPIRES the assumed relationship may be OR or AND, depending upon the file being searched.

Data base partitioning

This relates to the capability of partitioning the data base, for search purposes, so that a particular strategy can be applied only to a limited portion of the data base (e.g., specific years, specific types of documents). A somewhat related capability, possessed by STAIRS, is the ability to concatenate data bases. For example, the STAIRS user has the option of specifying that his search strategy is to be applied across several data bases existing in the system.

Sequential searching

This refers to the capability of being able to search on portions of the record for which inverted files have not been created. Typically, a system may conduct most of its searches on inverted files but, once a particular set of documents has been identified in this way, the system will allow the sequential searching of these records (e.g., a title scan). In RIQS all searching is sequential.

Profile storage

This feature allows a user to store a search strategy within the system. This pre-stored strategy can be called up and displayed or used to search against full or partial data bases contained within the system. The feature is most valuable, of course, in providing the capability for SDI (selective dissemination of information): the user can store a search strategy, representing his profile of current interests, within the system and simply visit the terminal, periodically (say monthly) in order to find references to relevant items added to the system since last he consulted it.

DISPLAY AND SECONDARY SUPPORT FEATURES

The major features in this category are shown in Table 2, which also indicates which systems possess which features. Again, the "minimal features" are denoted by an asterisk.

Search review

This feature, regarded as a minimal feature by participants, allows the user to review the status of his search. In response to some type of "recap" command the system will display the sets currently active, the number of citations each contains, and the strategy that caused each set to be created. The feature is obviously not needed in a system such as LEADER, which does not operate on request sets in the conventional sense.

Predefined formats

This also was regarded as a minimal feature by most participants. The feature allows the user to specify which of several standard formats for document records he would like to have displayed (e.g., title plus source, full citation plus index terms, title plus abstract). This feature is under user control only to the extent that he can select any of several possible formats. He may not, however, specify new formats (i.e., formats that the system has not predefined). The discussion on this feature pointed out that the slower the terminal in use the greater the need for abbreviated forms of output.

Field specification

This is related to the previous feature but places the output options more directly under user control. That is, systems having this feature allow the user to request a record format consisting of elements that he has specified. The record is built up according to his specifications and is not predefined by the system. For example, the user can request a display consisting of report number, title and abstract or author, title and abstract.

Rapid scan

This capability was regarded as being non-essential by most participants, and it is a capability that only a few systems claim. The feature allows the display of brief records in rapid succession, without the need for the user to hit a key to select on a record-by-record basis. The display continues until interrupted by the searcher.

Highlighting

The developers of the Data Central system feel that this is a very important feature. Most other participants said it was useful but by no means essential. It relates to the capability, in a text searching system, of indicating where in a record the words causing an item to be retrieved appear. In Data Central many techniques have been tried. Perhaps the most impressive makes use of a color terminal with the highlighted words appearing in a color different from that of the remaining text. In black-and-white terminals other techniques are possible, including the use of arrows, asterisks, blinking, variable intensity, and the dropping of the highlighted terms below the level of the others on a line.

Expanding

This relates to the capability of "expanding" the length of a document record displayed. For example, the searcher could first request title, then full citation, and finally abstract or index terms. In certain types of systems he may be able to "expand" to full text, either in digital or microfiche form. The need for an expand capability is probably more critical for slower terminals. That is, for a fast terminal it requires very little more time to display an abstract than to display citations. In fact, a really fast CRT display can probably show abbreviated versions faster than the user can scan them.

Sorting

This relates to the capability of sorting the set of records that match a particular search strategy and displaying or printing these records in an order specified by the searcher. Possible sorting options include author, journal title, and publication date (in direct or reverse chronological order). As seen in Table 2, few of the systems presently offer alternative sorting capabilities, but several systems are in the process of giving this capability to the on-line user.

Ranking

This is related to sorting. It refers to the ability of the system to present records to the on-line user in order of the degree to which they match his search strategy. Hopefully, a sequence by "degree of match" will also approximate a sequence of probable relevance. The LEADER system, by its very nature, generates a ranked output. STAIRS and ORBIT have limited ranking capabilities. Five possible ranking algorithms exist in STAIRS, including one based on the absolute frequency with which a search term appears in a document, one based on the frequency of occurrence of a search term in the corpus as a whole, and one based on word proximity within documents. I personally feel strongly that a ranking capability is needed in on-line, interactive systems, but the majority of the participants at this workshop did not agree with me, perhaps because their own systems do not offer this feature.

Computing

Three systems offer the user some computing facility. An example of this feature is a system which would allow the user to retrieve descriptions of all contracts worth in excess of, say, $10,000, and then do various arithmetic manipulations on the financial data (e.g., sum the values or average them). BASIS-70 offers some limited computing feature. Both RIQS and SPIRES interface with statistical computation programs and can offer the on-line searcher fairly sophisticated capabilities for computation.

Microfiche

This relates to the capability of being able to access a remote microfiche file from a searching terminal. File addressing is under control of the same computer that controls bibliographic searching. The user, once he has identified citations of interest to him through the bibliographic searching system, has the ability to request that microimages of the item be displayed at his terminal. The images may be displayed on the CRT that the search is conducted on, or they may be displayed on an adjacent tube. INTREX has the most sophisticated microfiche interface of this type. BASIS-70 also has the capability of interfacing with a microfiche store. DIALOG claims the capability but does not use it. Most participants agreed that a microfiche interface would be a highly desirable feature but that the cost of such interfaces at the present time make the feature hard to justify for most applications.

Offline printing

It was generally agreed that all on-line systems should give the user the ability to request than an off-line printout be made of all records satisfying a particular strategy. All eleven systems have this capability. Generally the results are mailed to the user at a later time.

Display of graphs

Both BASIS-70 and RIQS include the capability of presenting data in the form of graphs. In the case of RIQS quite sophisticated plots are made for various characteristics of searches that have been conducted (e.g., a typical plot would show how searches are distributed in terms of the amount of time consumed by each).

Batch retrieval

This relates to the capability of entering a search strategy at an on-line terminal, testing it out there, and then requesting that the strategy be used to search some data base in a batch mode. Typically, a very large system will maintain only a portion of its data base (perhaps the last two years) on-line. The remainder of the data base can be accessed only in an off-line, batch mode. The searcher should, however, be given the ability to have his on-line strategy used intact in a search of the off-line file.

Random citation selection

This feature allows the searcher to specify that he would like to see, from the complete set of documents that match his search strategy, a few items selected in a quasi-random manner. Several systems have the ability to generate such a quasi-random subset. For example, the ORBIT searcher can request that, from x citations that satisfy his search requirement, the system is to print 2, skip 10, print 2, skip 10, print 2.

INSTRUCTIONAL AND DIAGNOSTIC FEATURES

I chaired this particular session of the workshop. Below is a summary of my introduction and the framework for discussion that I attempted to establish.

The session deals with two distinct but related topics: (a) training of users, and (b) system diagnostic features. Training of users involves two phases: (1) initial training (i.e., giving enough training to allow a person to make some use of the system) and (2) "continuing education" (i.e., methods of raising the performance of the searcher, showing him how to use more sophisticated techniques and approaches, and informing him of system changes, including new capabilities).

As in most other areas of endeavor, there is no real substitute for hands-on experience. That is, people will learn best how to use on-line retrieval systems by practice. However, the new user needs a limited amount of instruction before he can ever get onto the system, even if this instruction merely tells him how to log-on. In considering the possibilities for training the new user, it is well to remember that two types of user may be involved (i.e., librarians or other information specialists on the one hand, and "end users" on the other) and that methods suitable for the training of one type may not be suitable or feasible to use in the training of the other.

It appears that there are at least four possible approaches to initial training:

1. Personal instruction. This involves a one-to-one relationship between the instructor and the "student." The student learns by using the terminal with the instructor at his shoulder. The

situation is probably ideal in many ways. Unfortunately it is very expensive and therefore not suitable for training of large numbers of potential users. It has other limitations too. If potential users are widespread, it is unlikely to be feasible to instruct them in this way. It is most likely to be feasible in the training of a few users in a central location.

2. Classroom instruction. This involves a one-to-many relationship between the instructor and the potential users. A classroom session, or series of classroom sessions, including demonstrations, is followed by periods allowing hands-on experience. The approach is less personalized than the one-to-one relationship but is also less expensive and more practical for the training of large numbers of users. It retains the advantage of providing a "live" instructor to answer questions and to assist with problems encountered in the training sessions. The classroom approach has been used successfully by NLM in the training of MEDLINE search analysts and has been used successfully by Informatics in training users of TOXICON, an application of RECON.

3. Printed instructional materials. There are at least three types of these: (a) user manuals, (b) summaries of basic capabilities, and (c) on-line printed instruction.

A complete user manual, describing all system features in some detail, is certainly needed. However, such a manual is more useful as a reference tool than as a training aid. The typical user manual is much too detailed to be suitable for use in the training of a new user; it gives him more than he can possibly absorb and more than he really needs as a beginning user. If a user manual is intended to be used for instruction, it needs to be clearly divided into easily digestible sections or lessons. The first section or lesson merely tells how to log-on to the system and presents only enough commands and capabilities to allow a user to conduct a relatively simple search. When he has mastered this, and proved it at the terminal, he moves on to the next lesson, which brings in new commands and more sophisticated procedures. He proceeds in this way until he has mastered the full system capabilities.

An improvement on the conventional user manual, for training purposes, is a brochure summarizing basic system capabilities and including, perhaps, a sample search. A number of such publications exist for various systems, including DATA CENTRAL and DIALOG.

On-line instruction is offered by several systems, including ORBIT, DATA CENTRAL and INTREX. In the MEDLINE application of ORBIT, for example, once the user has logged in he is asked to identify himself as a new or experienced searcher. The "new" searcher is given the opportunity to receive instruction in how to use the system at the on-line terminal. The instruction given may be in abbreviated or full form. DATA CENTRAL includes some quite sophisticated instructional capabilities, designed for CRT presentation, and these take the user through sample searches.

The subject of on-line instruction generated some discussion. Those systems that provide such facilities are strongly in favor of them, while those who do not are against the use of on-line connect time for training purposes. Their attitude is "why use valuable on-line time to tell the user something he can just as easily read in a printed manual?" This viewpoint has some justification. Unfortunately, many people are very reluctant to read printed instructional materials, and are more likely to accept the on-line tutorial because of its novelty. The investigators of Project INTREX discovered that users preferred on-line instruction, even when the on-line instruction was identical to instruction available in printed form. Moreover, a well designed on-line tutorial, especially one using a CRT, can be much more effective than a printed tool in giving the user "the feel" of the on-line system. This is particularly true when the tutorial takes the user through one or more sample searches, as in the case of the DATA CENTRAL system.

4. Audiovisual instruction. This form of instruction is probably the one most neglected. A few script-slide presentations exist (e.g., on DIALOG) but they are limited in scope. Script-slide presentations are essentially static and a more dynamic approach is clearly desirable. One possibility which seems to have considerable promise is the use of videotape. A well-designed videotape might prove to be a reasonably good substitute for personal instruction. The videotape would be used to show an experienced searcher at work at the terminal. The viewer would be looking over the shoulder of this searcher. With audio commentary, the searcher would explain how to log-on, how to conduct a simple search, how to display results, and so on. A second videotape could be used at a later time to introduce more sophisticated searching techniques. As far as I am aware, no one is using videotape for training in how to use on-line bibliographic systems.

The continuing education of the user presents different problems and challenges. There are also many more facets to continuing education. Within "continuing education" we can legitimately consider all forms of help given to the user by the system, all built-in tutorial features, and perhaps all the searching aids that the system provides. The object of this aspect of training is to increase the level of the user's performance, to help him when he goes astray, and to inform him of new system capabilities. Some possible "continuing education" features are:

(a) EXPLAIN. Several systems have a feature of this type. That is, they provide an "explain" command by which the user can obtain (on demand) explanations of system commands, error messages, and other system features.

(b) HELP. Help to the on-line searcher who finds himself in difficulties can be provided in several ways. First, in some systems (e.g., ORBIT) the command HELP will bring to the user any one of several pre-stored "solutions" to problems commonly encountered. The user is first presented with a list of commonly encountered problems. He selects from this list the problem that is applicable to his own situation and the system provides him with a generalized statement on how to proceed. This type of help is reasonably satisfactory as long as it is possible to identify the major problems likely to occur in the system, to label these problems in a way that allows the user to recognize them, and to present solutions that are clear to the terminal user. Obviously, it is not possible to anticipate all types of problems that might occur and treat them in this standardized form.

A better approach is to use the HELP command to bring in an experienced searcher at another terminal. Communication between experienced and inexperienced searchers is by way of these terminals. This is more personalized but it is also more expensive. It requires that a trained searcher be at an on-line monitor at all times, to await HELP messages and handle them accordingly. In this situation much searcher time is spent in waiting and it is unlikely that he or she could be very productive on other things while "on call" in this way. A second terminal has to be dedicated to the "help" mode if this method of offering aid is adopted.

It was generally agreed that the best approach to "help" is to provide it in live form by giving the terminal user a special telephone number to use when assistance is needed. The telephone number puts him in touch with a fully experienced searcher, who can help him with his specific problem. It was felt by most participants that this approach is much better than the on-line communication for two reasons: (a) it is easier to communicate and to identify the problem, and its solution, by telephone, and (b) most users prefer to deal directly with another human rather than through the medium of the machine.

(c) More advanced and sophisticated system techniques. The novice user should have some opportunity of learning more sophisticated procedures, either by moving to a more advanced "lesson" in an instruction manual or by calling up a more advanced tutorial at the terminal.

(d) New system features. The system must provide some mechanism whereby its users are kept current with new developments and new system capabilities. One method of achieving this, used by MEDLINE among other systems, is to have a NEWS command. When the user enters this command he is informed of recent changes in the data base, in the command language, in the error messages, and so on. Several systems also publish a newsletter to keep users informed of new developments of this type.

(e) Other aids to the searcher. Although not strictly instructional, a related set of system features are designed to make the system easier for the searcher to use. "Ease of use" features include features designed to reduce the likelihood of error and features designed to reduce the amount of keying that needs to be done. Features of this type include: display of "menus" from which the user makes a selection; acceptance of abbreviations, including abbreviations for command names; acceptance of common misspellings, including spelling errors in command names (BASIS-70, for example, will accept common variants) and index terms (RECON and DIALOG accept spelling errors in index terms in the sense that they display terms that are alphabetically close to the misspelled term); and the RENAME command (as in ORBIT) which allows a user to change command names, if he wishes, to a form with which he is more familiar (e.g., because he uses these names in another system).

In concluding my introduction to the subject of instructional features I raised the following additional points:

1. That instruction of users of on-line systems involves three distinct aspects: (a) mechanics of using the terminal, (b) construction of logically sound searching strategies, and (c) internal "intellectual" factors, largely associated with the characteristics of the particular data base involved. I suggested that these aspects, from the viewpoint of instruction, are in order of increasing complexity. It is relatively easy to train people in how to use the terminal (i.e., pure mechanics). It is somewhat more difficult to teach them how to reduce their search requirements to the form of a Boolean strategy. Probably most difficult, however, is the problem of informing users on the characteristics of the data base. This is particularly true where the system is based upon a large controlled vocabulary of terms assigned to documents by trained indexers. To use the system most effectively the searcher needs to know something of indexing policies and protocols. He also needs to understand the controlled vocabulary and be able to find his way around it. It is not easy to impart this information to the user. It takes several weeks, for example, to train an indexer at the National Library of Medicine. Clearly, the biomedical practitioner using MEDLINE cannot learn all the nuances of indexing and system vocabulary. It is for this

reason that systems based on uncontrolled vocabularies (i.e., natural language) are likely to be used more effectively by the person who is not an information specialist. Training these users in how to exploit natural language systems is also likely to be easier than training them how to use controlled vocabulary systems effectively.

2. Possibilities for training on-line are limited by the type of terminal in use. A CRT or other form of video display seems to offer much greater scope for an imaginative, dynamic instructional approach than a typewriter terminal does.

3. We may not have fully explored the possibilities for applying techniques of CAI (computer-aided instruction) in training users. Quite sophisticated CAI programs exist in many other areas of endeavor. Such techniques could be used either in initial instruction of the potential user or in actually "taking him by the hand" and leading him through the conduct of a search. Some work on CAI techniques applied to the training of users of on-line retrieval systems has been carried out by Caruso at the University of Pittsburgh and by Lyman, working with the PLATO system at the University of Illinois.

4. A question worth raising is "Should we be spending more time and effort in seeking techniques to make on-line retrieval systems easier to use?" For example, do we need to make such systems less sensitive to simple errors of spacing, punctuation, and spelling? Or should we adopt the attitude that present systems are adequate, from a human factors standpoint, and that users should be required to be absolutely accurate? This appears to be a somewhat controversial question. Some believe that present systems are easy enough to use, even by the person who is not an information or data processing

specialist, while others believe that on-line systems must be made simpler and more "forgiving" if they are to be used extensively by scientists and other professionals.

Some major instructional and diagnostic features of the eleven systems are shown in Table 3. Most are self-explanatory.

Manuals

All eleven systems provide some form of user manual to explain (1) how to use a terminal, (2) how to log on and off the retrieval system, (3) what commands are available and how they work, (4) what data bases are accessible and how they are structured, (5) how to develop a good search strategy, and (6) what to do when X happens. These manuals are of variable quality. Some are complete and quite readable and others are incomplete and/or virtually unreadable.

Modes of Instruction

The major approach(es) taken when educating the first-time user.

- SYSTEM - the system is programmed to help in the training
- CLASS - live or filmed courses are given periodically or upon request
- PERSONAL - an expert sits down with the user at the terminal
- READING - the manual is read

On-Line Training

The user may use the terminal itself to obtain information concerning (1) terminal and typing problems, (2) the command repertoire, (3) characteristics of the data base, (4) hints for good searching, (5) common pitfalls and their remedies.

ORBIT	- The beginning user easily falls into this material.
DATA CENTRAL	- The beginning user easily falls into this material. When in verbose mode, system responses cue the user.
SPIRES	- The beginning user easily falls into this material. He can go through a sample dialog annotated with explanations.
RIQS	- The knowledgeable user can attach a tutor to the retrieval system. The tutor is the unique source of command and error explanations.
INTREX	- When in verbose mode, system responses cue the user.
LEADER	- When in verbose mode (the only mode), system responses cue the user.

Data Base Overview

Information concerning (1) the size of the data base, (2) the types of material in it, (3) the time span covered by the material, (4) the fields making up a record, (5) the searchable fields, and (6) special strategies to use or pitfalls to avoid with the data base. This type of information is available on-line in the systems checked in this column.

Sample Searches

Three of the systems allow the storage of searching strategies. The user, if he knows of the existence of a prestored strategy, can call it up and incorporate it into his own search formulation. These prestored strategies can include search logic. Unfortunately, the searcher must know of the existence of these strategies in order to be able to use them. If he enters a particular term that is the name of a prestored strategy, he is not automatically notified of its existence.

On-Line Documentation

"One page" descriptions of all commands and error messages, available on-line. For example, the command EXPLAIN in ORBIT will generate this type of description.

Search Logic Tracing

The searcher can request a detailed description of how his multi-part search request led to the number of hits reported.

Live Help

Either a telephone number to call if desperate or a message command for requesting help from the on-line human consultant.

Vest Pocket Card

A durable folder containing command names and an explanation of how to get complete command descriptions. May be a card attached directly to the terminal.

Diagnostic Features

Diagnostic features are incorporated into the system to determine failures and problems occurring, and thus to suggest possible corrective action. The only real way to obtain detailed information on user successes and failures, and factors affecting performance of the system, is by a fairly extensive evaluation based on a sample of actual searches. We were not particularly concerned with this type of one-time evaluation at this meeting. Rather, we were concerned with procedures that might be used to analyze the system routinely on a continuous basis.

One approach is the on-line questionnaire that the user completes at the end of his search. Such a questionnaire can be used to determine various characteristics of the user, purpose of the search, and the user's subjective assessment of its value to him. This, coupled with the connect time for the search, gives some general idea of its characteristics. The on-line questionnaire has some value in revealing who the users are, for what purpose they use the system, and their degree of satisfaction with the search results.

Both MEDLINE and RECON have made use of on-line questionnaires of this type, but both systems have since abandoned this feature. In the case of RECON this decision was made because it was found that very few users completed the questionnaire anyway.

Another possible diagnostic element is the "comments" feature, whereby a user can make suggestions for improvements to the system or record problems he has had in using it, by typing these comments at the terminal at the time the problem arises or the idea springs to his mind. These messages are analyzed by the system managers at some later time. As seen from Table 3, several of the eleven systems have this feature.

By far the most useful "diagnostic" feature, however, is the capability for on-line monitoring of what users are doing. Several of the systems do conduct various monitoring operations and summarize the results in a "monitor log" (Table 3). With this type of on-line monitoring of use we are given the possibility of learning more than ever before about how scientists and others use information retrieval systems, and with what degree of success. Unfortunately, monitoring of on-line users raises certain ethical and possibly even legal problems. Monitoring of a user without his express permission might possibly be construed as a form of "wiretapping," an invasion of privacy to which the United States is particularly sensitive at the present time.

Several of the participants at this meeting expressed the feeling that the user must be told that he is being monitored and perhaps given the option to cut off the monitoring operation. In SPIRES, for example, there is a command available that will avoid the monitoring operation, although few users know of its existence. Even if the user gives his permission, of course, we are faced with the problem that the user who knows he is being monitored may not behave in exactly the same way that he would if he were unaware of the monitoring operation.

Clearly, a system should be allowed to do certain types of monitoring, even if the monitoring is restricted to purely statistical summaries. The question is:

"What, if anything, can legitimately be monitored and what cannot?" It seems clear that the system should be allowed to monitor, record and analyze statistical aspects of use. Such statistical data will include number of uses, number of uses per terminal, distribution of search sessions by connect time and by CPU time expended, number of simultaneous users, and so on. Data of this type are actually reduced to the form of graphs, which can be displayed on a CRT, in the RIQS system. Where the on-line user is required to identify himself by any category (e.g., experienced versus inexperienced), or to identify his search by type or purpose, it seems legitimate to include these variables in any overall data collection activity.

In fact, it seems reasonable that the system should be allowed to monitor, and collect data on, the behavior of users in the aggregate, as opposed to collecting data on the individual, identifiable user. Such aggregate data would include statistics on how the system is used (e.g., data on frequency of term usage, frequency of command usage, and frequency with which particular sources are retrieved) and the types of problems encountered (e.g., frequency of use of various error messages, frequency of use of the HELP command, and the specific form of help requested, and the frequency and type of use made of the EXPLAIN command). This type of "aggregate" monitoring certainly seems justifiable and is of great importance to system managers in showing how the system is used and what might be done to improve its performance.

The system is on more dangerous ground, perhaps, if it monitors the individual user. Monitoring of an individual user, either by observing him directly from a second terminal or by recording his dialog with the system for later analysis, would certainly have potential value. A detailed analysis of steps used in searches could identify the types of problems that appear to be most prevalent and could indicate which system aids are used successfully and which are little used or used poorly. From this type of analysis, at the level of the individual search strategy, a great deal could be learned that might be used in the later improvement of the system itself or in the improvement of methods of training users.

Unfortunately, it is this type of monitoring that is most likely to be construed as an invasion of privacy. The problem is likely to be more important in certain situations than in others. In an academic environment monitoring might be regarded as less objectionable than it would be in an industrial environment. Users of commercial services, who are paying for use of the system, may be more opposed to such monitoring than users of non-commercial services, especially a service provided to the user at no direct cost to him.

In many systems, however, the terminal is identifiable but the user himself is not. For example, the log-in operation in MEDLINE identifies the user organization only. Many individual users may search the system from a particular terminal during the course of a day. These users are essentially anonymous because the system does not require them to identify themselves by name (unless they happen to request an off-line printout). It is difficult to see why an anonymous user should object to being monitored in many situations. The industrial situation, where a user may not want anyone else to know the subject matter of his current interests, is a different situation.

It was reported by the representative of LEADER that this system routinely monitors users at Lehigh University and has even gone as far as to contact users and make suggestions as to how they might improve their use of the system, based on staff analysis of search approaches used. No objections to these activities have been raised by users at Lehigh.

It would seem desirable, however, for an operating system to obtain some blanket approval of its monitoring operations. In some situations, for example, it should be possible to include a "permission to monitor" in any agreement made between the system and a particular organizational user.

SUMMARY

I believe this workshop was extremely valuable in bringing together, for constructive discussions, representatives of most of the leading on-line bibliographic systems in use in the United States in 1973. Although several of these systems are commercial competitors, the meetings were held in an atmosphere of mutual respect and a considerable amount of valuable communication was achieved. A remarkable level of agreement was reached on the desirable features and characteristics of on-line systems of this type.

If the workshop had any weak points, it was the fact that it concentrated largely on existing systems and present capabilities. Very little time was devoted to consideration of future systems and future capabilities. In the summary session, chaired by Parker, possible future trends were considered, but not in any great depth. The major thrust of this discussion was in the area of system interfaces. It was generally agreed that on-line information retrieval systems should not be regarded as completely independent entities. They will be required in the future to interface with other systems. Such interfaces will be with other bibliographic systems, raising problems of compatibility and convertibility of vocabularies and searching strategies, and with other types of systems (e.g., interfaces with statistical packages, with text editing systems, with photocomposition systems, with CIM and COM systems). System interface design may be one of the most challenging problems facing us in the years to come.

APPENDIX 1

PANEL OF EXPERTS

Professor Jack Belzer KAS Center University of Pittsburgh 135 North Bellfield Street Pittsburgh, Pennsylvania 15213	(412) 621-3500 x6352
Mr. John L. Bennett Research Staff Member IBM Monterey & Cottle Roads San Jose, California 95114	(408) 227-7100 x2682
Professor F. Wilfrid Lancaster Graduate School of Library Science University of Illinois Urbana, Illinois 61801	(217) 333-7742
Professor Richard E. Nance Computer Sciences Center Institute of Technology Southern Methodist University Dallas, Texas 75222	(214) 363-5611 x2231
Mr. Donald E. Walker Artificial Intelligence Group K 2096 Stanford Research Institute Menlo Park, California 94024	(415) 326-6200 x3071

APPENDIX 2

SYSTEM SNAPSHOTS

ORBIT

1. large bibliographic data bases (ERIC, CHEMCON, NLM)
2. no full text search capability at present
3. on the Tymshare network
4. used primarily by intermediaries or intermediary-end user teams
5. primarily accessed by non-CRT terminals
6. acts as a service center as well as selling software
7. terminal tends to be in a central location
8. commercial rates
9. PL-1, with some assembly language
10. 360/50 and 370/155

DIALOG

1. large bibliographic data bases (NTIS, ERIC, Pandex)
2. full text search capability
3. on the Tymshare network
4. used primarily by intermediaries or intermediary-end user teams
5. primarily accessed by CRT terminals but some use of teletypes

6. acts primarily as a service center

7. terminal tends to be in a central location

8. commercial rates

RECON

1. large bibliographic data bases (TOXICON, NASA) and management information data bases

2. full text search capability

3. on the Tymshare network

4. used primarily by intermediaries or the end user

5. primarily accessed by non-CRT terminals

6. acts as a service center as well as selling software

7. terminal tends to be in a central location but can be user's personal terminal

8. commercial rates

9. assembly language

10. 360/370 series, models 40, 50, 65, 155

STAIRS

1. large textual data bases (Eng Index, Suny BCN, Congress)

2. designed primarily for full text

3. international access by IBM for in-house use

4. used both by intermediaries and end users

5. primarily accessed by CRT terminals; some teletype

6. intent is to sell hardware not service center

7. terminal tends to be in a central location but can be user's personal terminal

8. outside IBM charging is up to the buyer

9. assembly language

10. 360 and 370 series

DATA CENTRAL

1. large textual data bases (EPA) and management information data bases

2. designed primarily for full text

3. users access system remotely

4. used primarily by end users but with some data bases has been used by intermediaries

5. primarily accessed by CRT terminals

6. intent is to lease software, not service

7. terminals tend to be personal

8. commercial rates

9. primarily assembly language; some COBOL

10. 360/40 and larger configurations

BASIS-70

1. large textual data bases (NTIS, CHEMCON) and management information data bases
2. no full text capability at present
3. on the Tymshare network
4. used primarily by intermediaries or end user
5. primarily accessed by CRT terminals (3000 baud)
6. intended for service to in-house intermediaries
7. terminals tend to be personal
8. outside the institute, commercial rates apply
9. primarily FORTRAN; some CDC assembly language
10. CDC 6000 series

LEADER

1. large bibliographic data bases (Eng Index, Chemical Condensates)
2. no full text search capability in the conventional sense
3. used by both on-campus and off-campus users
4. used primarily by end users or end-user intermediary teams
5. primarily accessed by CRT terminals
6. acts as a service center

7. terminals tend to be in a central location

8. outside the university commercial rates apply

9. FORTRAN

10. CDC 6400

SPIRES

1. primarily management information data bases but one large bibliographic data base (MARC)

2. no full text search capability

3. use restricted to terminals on Stanford campus at present, although it is possible to access remotely

4. used primarily by end users

5. primarily accessed by non-CRT terminals

6. intent is in-house service although both distributing software and considering service center role

7. terminals split between personal and in a central location

8. minimal charges

9.

10. 360/67 and 360/91

NASIS

1. large photographic description data base (ERTS) and management information data base

2. full text search is possible

3. primary access is via the federal telephone network

4. used both by end users or intermediaries

5. primarily accessed by non-CRT terminals; can be accessed by CRT terminals up to 1200 baud

6. acts as a service center

7. terminals tend to be personal

8. no charging algorithm

9. PL-1

10. 360/50 is minimum configuration needed

RIQS

1. personal data bases -- both bibliographic and management information

2. there are no inverted indexes and no full text search

3. use restricted to faculty and staff at Northwestern University

4. used primarily by end users

5. primarily accessed by non-CRT terminals

6. intended for in-house searching

7. terminals are both in a central location and in personal locations

8. no charging algorithm

9.

10. CDC 6400

INTREX

(in transition from experimental to operational status)

1. small bibliographic data base (INTREX)

2. full text search capability exists

3. use restricted to MIT campus at present, but can be accessed remotely

4. used primarily by end users or end user-intermediary teams

5. primarily accessed by CRT terminals

6. intended for in-house searching

7. terminals are in a central location but can be personal

8. no charging algorithm

9.

10. at present being reprogrammed for 370/165

APPENDIX 3

SYSTEM REPRESENTATIVES

Dr. Louis Stern — (215) 867-5071 x323
Center for the Information Sciences
Lehigh University
Bethelehem, Pennsylvania 18015

Mr. Larry Stevens — (301) 770-3000 x217
Informatics, Inc.
6000 Executive Boulevard
Rockville, Maryland 20852

Mr. Richard Giering — (513) 426-3111
Mead Technology Laboratories
Research Park
Dayton, Ohio 45432

Dr. Richard Marcus — (617) 253-1000 x2340
Project Intrex 35-406
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Dr. Charles Goldstein — (216) 433-4000 x6660
Computerized Information Systems Office
National Aeronautics and Space Administration
Lewis Research Center
Cleveland, Ohio 44135

Dr. Roger Summit — (415) 493-4411 x45034
Mr. Mark Radwin — " x45769
Lockheed Palo Alto Research Laboratory
Palo Alto, California 94304

Mr. Donald Black — (213) 393-9411 x7513
System Development Corporation
2500 Colorado Avenue
Santa Monica, California 90406

Mr. Dave Colombo (614) 299-3151 x3240
Battelle Memorial Institute
505 King Avenue
Columbus, Ohio 43201

Mr. Stan Friedman (914) 765-2123
International Business Machines Corp.
Armonk, New York 10504

Mr. Larry Rosen (415) 321-2300 x4531
Stanford Computation Center
116 Polya Hall
Stanford University
Stanford, California 94305

Dr. Benjamin Mittman (312) 492-3682
Mr. Wayne Dominick
Vogelback Computing Center
Northwestern University
2129 Sheridan Road
Evanston, Illinois 60201

VISITORS

Dr. Dennis Fife (202) EM 3-4040
National Bureau of Standards
Connecticut Ave. & Van Ness, N.W.
Washington, D. C.

Mr. Jim DeiRossi (202) EM 3-4040
National Bureau of Standards
Connecticut Ave. & Van Ness, N.W.
Washington, D. C.

Mr. Richard Lee (202) 632-5818
Office of Science Information Service
National Science Foundation
1800 G Street, N.W.
Washington, D. C. 20550

Prof. David Thompson (415) 321-2300 x4474
Industrial Engineering Department
Stanford University
Stanford, California 94305

Table 1

SEARCH NEGOTIATION FEATURES

	REQUEST SETS *	DICTIONARY ACCESS *	RELATED TERM CAPABILITY	HIERARCHICAL THESAURUS	CAPABILITY OF INCORPORATING SYNONYM TABLES OR TERM HIERARCHIES	SEARCH FIELD CONTROL *	BOOLEAN OPERATORS *	WORD PROXIMITY OPERATORS	ARITHMETIC OPERATORS	SUFFIX REMOVAL *	PHRASE DECOMPOSITION	DATA BASE PARTITIONING	SEQUENTIAL SEARCHING	PROFILE SEARCHING
ORBIT	x	✓	✓	✓	✓	x	✓			✓		✓	✓	↑
DIALOG	x	✓	x	x	✓	✓	✓	✓		✓		✓		✓
RECON	x	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		↑
STAIRS	x	✓	✓		✓x	x	✓	✓	✓	✓			✓	✓
DATA CENTRAL	↑	✓			✓	x	✓	✓	✓	x ✓				
BASIS-70	x	✓	✓			✓	✓		✓				✓	↑
LEADER		x	x							x ✓	x			
SPIRES		✓	✓	✓			✓			✓	x	✓		✓
NASIS	x	✓				✓	✓		✓				✓	✓
RIQS	✓					✓	✓	x	✓	↑			✓	✓
INTREX	✓						✓	✓		x	x	✓	✓	

* minimal features x assumed automatically ✓ under user control

↑ the capability is being added to the system

DISPLAY AND SECONDARY SUPPORT FEATURES

	SEARCH REVIEW *	PREDEFINED FORMATS *	FIELD SPECIFICATION *	RAPID SCAN	HIGHLIGHTING	EXPANDING	SORTING	RANKING	COMPUTING	MICROFICHE	OFF-LINE PRINTING *	DISPLAY OF GRAPHS	BATCH RETRIEVAL	RANDOM CITATION SELECTION
ORBIT	✓	✓	✓			✓		✓			✓			✓
DIALOG	✓	✓	✓			✓	↑			✓	✓		✓	✓
RECON	✓	✓	✓	✓	✓	✓	✓				✓			✓
STAIRS	✓	✓	✓			✓	✓	✓			✓		✓	✓
DATA CENTRAL	✓	✓	✓		✓	✓	✓				✓			✓
BASIS-70	✓	✓	✓			✓	↑		✓	✓	✓	✓		
LEADER						✓		✓			✓		✓	✓
SPIRES		↑	✓	✓			↑		✓		✓			
NASIS	✓	✓	✓				↑				✓		✓	✓
RIQS	↑	✓	✓	✓		✓	↑		✓		✓	✓	✓	
INTREX	↑	✓	✓	✓	✓	✓				✓	✓			✓

* minimal features

↑ capability planned or now being added

Table 3

INSTRUCTIONAL AND DIAGNOSTIC FEATURES

	MANUALS (A*)	MODES OF INSTRUCTION: SYSTEM	CLASS	PERSONAL	READING	ON-LINE TRAINING	DATA BASE OVERVIEW	SAMPLE SEARCHES	ON-LINE DOCUMENTATION *	SEARCH LOGIC TRACING	LIVE HELP *	VEST POCKET CARD *	COMMENTS	MONITOR LOG
ORBIT	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓
DIALOG	✓		✓		✓				✓		✓	✓		✓
RECON	✓		✓		✓				✓		✓			✓
STAIRS	✓		✓					✓	✓	✓	✓	✓		
DATA CENTRAL	✓	✓				✓			✓		✓			
BASIS-70	✓		✓		✓				↑		✓			
LEADER	✓	✓		✓		✓			✓		✓		✓	✓
SPIRES	✓	✓	✓			✓	✓		✓		✓	✓	✓	✓
NASIS	✓		✓					✓	✓		✓		✓	
RIQS	✓			✓		✓	✓	✓	↑		✓		✓	✓
INTREX	✓	✓		✓		✓	✓		✓	✓	✓	✓	✓	✓

* minimal features are starred

↑ indicates that this feature is being added to the system at this time