DEVELOPMENT GRANT PROPOSAL

Project Title: SPIRES
(Stanford Physics Information Retrieval System)

Submitted by: Stanford University
Stanford, California

Initiated by: Edwin B. Parker
Associate Professor of Communication
Institute for Communication Research
Department of Communication
Stanford University

Telephone: Area Code 415
Number 321-2300
Extension 2755

Starting Date: June 1, 1968

Duration: 30 months

________________________
Edwin B. Parker
Principal Investigator

________________________
Clifford F. Weigle
Executive Head
Department of Communication

________________________
Research Administrator
Abstract

This proposal requests funds for the second phase (the development phase) of what has been planned as a five-year project. The primary goal of this phase is to develop an operational on-line reference-retrieval system that meets the needs of scientists at Stanford (including the Stanford Linear Accelerator Center). The first phase (the research phase) was supported by an NSF Grant for an 18-month period from December 1966 to May 1968. That phase provided a time for research, for planning, and for development of a prototype demonstration system. The first annual report, a copy of which is appended to this proposal, describes the progress during the first year. The expected accomplishments of the remaining months in phase one are outlined in a later section.

The development phase, lasting 30 months from June 1968 to November 1970, is intended as a period in which the on-line reference-retrieval system will be fully implemented and the behavior of the system and its users studied. On-line consoles (both typewriter and CRT display) and computer time would be provided during phase two to a group of physicists and librarians willing to use the system experimentally during its development and to permit the project staff to study their use of the system. During this period the reference-retrieval capability will be expanded to other areas of science beyond physics. The system would become generally available to any member of the Stanford community who has the funds to open an account with the Stanford Computation Center and to obtain access to an appropriate computer console. During this phase, planning would begin for other information services beyond the initial implementation of a reference-retrieval capability. A likely area for such development is that of on-line retrieval of particle cross-section data.

The third phase of the project, for which funds are not now requested, would be concerned with completing the transition from a development project to an economically viable reference-retrieval system to be provided as a service of the Stanford Library with the
assistance of the Stanford Computation Center. The third phase would also be concerned with continued research into information needs of users and with development of other information services beyond the reference-retrieval application.

**Phase One Plans**

Several tasks are expected to be completed by the end of the present funding period (phase one). During this period (ending May 1968), the present demonstration system is being modified to include the disk storage input and output routines. This is required to expand beyond the core-resident demonstration to a system viable with a large data base. Several modifications will be made in the demonstration programs -- to insert the hash coding procedure into the retrieval routines, to change the byte headings to bit headings, to expand the size of the block length, to add additional attributes to the input and data base building routines, and to program an update procedure permitting deletion and replacement of documents in the data base (at present, documents cannot be changed or deleted without regenerating the entire data base, although new documents can be easily added).

The production input of SLAC preprint document references and footnote citations should begin during this period, as well as the production input of *Nuclear Science Abstracts* (after completion of debugging and testing of the NSA input routines). The system capability should be expanded to add citation indexes and citation searches, and to add a personal file capability.

In addition to this programming, planning will continue for revision of the index structure implemented for the demonstration system, for provision of a generalized correction and update procedure, and for modifying the retrieval programs to run fully interactively with multiple terminals from a 100,000 byte partition in the 360/67 system at the Stanford Computation Center.
Phase Two Plans

There are two basic parts of the problem to be solved during phase two, the phase for which funds are being requested. One is the implementation of a fully interactive multiple-terminal reference-retrieval computer system. The other is detailed study of the information-seeking behavior of physicists in such an environment. The second goal is obviously dependent upon the first.

Two parts are necessary for provision of the "on-line" reference-retrieval capability. One is the set of systems programs for handling multiple terminals simultaneously with satisfactory response times. The other is the set of programs for the reference-retrieval application to run in that multiple-terminal environment. The present 2250 CRT display scope demonstration and the "remote batch" searches from 2741 typewriter terminals when modified during the remainder of phase one as outlined above, will provide the nucleus of the applications programs. With only slight modifications, this set of application programs can be made to run in a dedicated partition of 100,000 bytes of core storage on the 360/67, serving multiple terminals through the already operational Stanford Terminal Processor routines.

The Stanford Computation Center has agreed to provide the necessary interface routines to connect the dedicated partition containing the SPIRES system with the terminal driver routines. Specifications for such interface routines have been prepared. Within the SPIRES partition most lines of terminal input would be processed to completion before starting work on the input from the next terminal. When the program is waiting for input from disk storage, Central Processing Unit (CPU) cycles will be released to the batch jobs running in the background from another part of core, rather than trying to swap users within the SPIRES partition. The details of the procedure, which can be thought of as a special purpose "event-driven" (as distinguished from "clock-driven") time-sharing system, are presented in the appended copy of the SPIRES first annual report.
It is expected that multiple typewriter terminals (IBM 2741) can be supported for this application by this time-sharing system by September 1968, and that multiple CRT display terminals (IBM 2260) can be supported by December 1968, possibly earlier if the hardware delivery schedule can be speeded up. (The Stanford Computation Center has agreed to provide the software support for the 2260 CRT display terminals.)

The basic design of the Stanford 360/67 hardware configuration, system software, and reference-retrieval application programs seem particularly suited for on-line reference retrieval by a large number of users from a large data base. Given the present hardware and remote terminal processing routines, 62 terminals could be handled simultaneously. Further expansion is not precluded. The retrieval system is designed for efficient access to extremely large data files. Each index file (e.g., the author index) can be split into as many sections as necessary to permit each section to be brought into core with a single access to the disk. A "hash coding" address calculation from the query term (e.g., the author's name) will permit the appropriate section to be accessed immediately. To minimize index reorganization costs, it will be necessary to use "overflow" sections, requiring in some cases a second or even third disk access, but the average can be maintained at less than two disk accesses per query term, by periodic reorganization of the index files. The amount of processing to be done in core memory is minimal (considering the CPU rate of the 360/67) so disk access frequency is expected to be the limiting factor in providing adequate response time. The first system response to a query, which is the count of the number of documents meeting the given specification, can thus be provided quickly from information requiring not more than two disk accesses per query term. When output of document references is requested, one disk access per document is required, but the time required to print out the information at the terminal is long relative to the disk input/output speed. Therefore the system should be able to retrieve and format for output second and succeeding document references while the earlier ones are being communicated to the terminal.
As the system expands beyond the point where a one-level hash coding scheme is efficient, a two-level indexing scheme requiring one additional disk access per query term could be implemented. (The first level would be an index directory.) Alternately, the file can be partitioned (e.g., Nuclear Science Abstracts in one partition, SLAC preprints in another, etc.) with the user specifying which database he wished searched. For extremely large files (e.g., a university library science collection, including entries for each article in each journal), both strategies could be used. Actual response times under maximum terminal usage obviously cannot be calculated at this stage of development, but there seems to be no reason at present to suspect unsatisfactory response times.

Meanwhile, using primarily the remote batch capability for searching from 2741 terminals (with usually about two minute turn-around time), development and use of the reference-retrieval program itself can continue in parallel with development of the time-sharing system. A major feature, required by the second goal of the project, will be the programming of the monitoring, commenting, and questionnaire features of the system and the associated statistics collection and analysis routines. The intent is to be able to completely monitor all interaction with the system in order to study how users interact, what parts of the system are most used, and where users may have difficulties. The commenting feature will permit users to send messages to the SPIRES staff at any point during the search, whenever any suggestion occurs to them. The questionnaire feature will permit brief questionnaires to be administered on-line at the conclusion of some searches, to evaluate the relevance of the retrieved references (as perceived by the user) and to get a standard measure of user satisfaction with searches.

Several features need to be added to improve the quality of the reference retrieval. Two most important features needed are an optional "word stem" feature permitting searches for all title words with the same stem as the query term (e.g., scatter, scatters, scattering, scattered, etc.) and a title word synonym procedure in
which a previously constructed synonym dictionary is available in machine-readable form (as a result of prior analysis of the collection), permitting searches to retrieve titles containing words synonymous with the query term. Revision of the present index structure is planned to include as the first level of each index an easily-browsed dictionary of terms in that index, including the number of documents in the collection associated with that index term. The "personal file" capability should be expanded to provide a Selective Dissemination of Information (SDI) service, by posting document references to a temporary personal file for each participant in the system.

Use of the system by physicists will undoubtedly lead to several other areas in which improvement and expansion of the programming will be required to provide satisfactory service. For example, the query language may need to be revised into a "computer prompt and user answer" dialogue. Provision for weighted term searches also might be made. It will probably be useful to have a user's manual in programmed instruction form. Study of the interactions of users with the early version of the system will guide such later development. In addition to such system changes and modifications apparent to the user, it will undoubtedly prove efficient to reprogram some parts of the system to provide faster response time. Present versions of PL/1 object code run considerably less efficiently than assembly language code, for example, and substitution of assembly language subroutines for the frequently used sort and merge routines may prove useful.

The data base should be expanded beyond SLAC preprints and Nuclear Science Abstracts during phase two. For the physicists, the most useful data base to add will be the DESY Index produced (in machine-readable form) at the DESY high energy physics labs at Hamburg, Germany. Preliminary discussions have been held with DESY and no difficulties in getting magnetic tape copies of their index are foreseen. This index has the advantage of associating index terms with preprint and report documents. There is a noticeable overlap with the SLAC preprint collection, but the addition of DESY indexing terms makes that feature an asset.
The data base should also be expanded to include good coverage of the physics journal literature, accessible via citation indexing. One possible source of such data for selected physics journals, keyboarded at National Science Foundation expense, is the data base associated with M. M. Kessler's Technical Information Project (TIP) at MIT. Informal conversations with the American Institute of Physics have indicated that this data base is likely to be made available on a continuing basis through AIP. Continued availability of this data base, although desirable, is not essential to the success of the project.

A primary consideration in expansion of the data base is the population to be served. The expenses of development of the system are virtually the same whether the population of users is about 100 (the Stanford physics population) or more than 1000. The additional costs associated with the larger user population (in an environment where computer consoles are readily available on campus for other purposes) are a computer storage cost (for a larger data base, for example, in areas other than physics), a data acquisition cost, and the increased computer time charges associated with use of the system. There are two primary reasons why a physics information system project should concern itself with such a larger population of users. One is that to make the system ultimately viable economically (and thereby regularly available for physicists) a broader economic base is desirable. The other is that our interviews have confirmed the common sense notion that literature searches are both harder and more important (relative to informal communication) when the information problem is in a related area rather than a core physics problem. The advantages to physicists of expanding to a more general scientific reference-retrieval system seem sufficient to justify the additional expense (particularly if utilization costs for computer searches by other than physicists are not subsidized). Therefore we propose to include the Institute for Scientific Information Citation Index (in its machine-readable rather than hard-copy form) as a journal literature data base, providing coverage of some 1600 scientific journals.
When the additional advantages of making such an on-line reference-retrieval system available to the general population of scientists at Stanford are considered, the advantages seem compelling. The SPIRES project is working closely with the Stanford Library Automation Project (Project BALLOTS) to integrate the operating system that SPIRES develops into the total library operation. There is close coordination between the two projects in order to avoid any overlap of effort. SPIRES is concerned primarily with development of user services, whereas BALLOTS is concerned primarily with internal library operations. Expansion of the SPIRES data base to include non-physics scientific and technical journal references should make it easier to develop SPIRES to the point where it becomes an important user interface with the Stanford libraries.

The general plan for phase two, then, is to develop the physics reference-retrieval system in such a way that there is maximum compatibility across disciplines to make it possible for other than physicists to benefit early from the system development. By the end of phase two, the reference-retrieval system should look more like a general campus information utility than merely a specific system for physicists. However, the center of focus remains on the physicists who are the target population for development and detailed systematic research. Information gained from this study of physicists should make it easier to develop within phase two a general "retailer" science information system at Stanford. This system should then provide the means through which the Stanford community can obtain access to the information in the developing national information systems of different scientific disciplines (which presumably are concentrating first on the "wholesaler" aspects more appropriate to their national missions).

As an additional part of phase two, plans will be made for later expansion of the information system beyond the reference-retrieval system, which is viewed as the first major segment of a more general program for meeting scientific information needs. For purposes of this planning, physicists will remain the center of focus. It seems likely at this early point that nuclear particle data will be the
most valuable feature to add next to such a general physics information system. Addition of abstracts as well as document references will be highly desirable, but will depend on mass storage costs and the availability of machine-readable text.

**Phase Three Plans**

By the end of phase two a technically viable reference-retrieval system should be fully operational. The University's plan is to have that operational system taken over and administered by the Stanford Library (using the computer facilities of the Stanford Computation Center) as a logical user-service extension of the on-line automation of internal library processes, including acquisitions, cataloging, and circulation control. (The Stanford Library Automation project is one of the three major library automation projects in the country and expects to be one of the earliest major libraries involved in this kind of operation.) There would be four elements of cost associated with operations in phase three: (1) hardware costs associated with operation of a core-resident system, (2) computer utilization costs, (3) personnel costs associated with system maintenance (including one systems programmer), (4) costs of acquisition and storage of the data base. It is hoped that the utilization costs will be provided by users (including the Stanford Library in its operations) at a level sufficient to justify both hardware and utilization costs. (One of the objectives of the Library Automation project is to study the economic factors involved in an on-line library system with associated on-line information services.) It is anticipated that a subsidy will be requested of the National Science Foundation during the initial operational stage (phase three in the present planning) to cover the personnel costs associated with system maintenance and the costs of acquisition and storage of the data base. It is currently estimated that the requested subsidy will be at the level of approximately one-third of the development costs during the final year of phase two. (The present proposal requests funds for phase two only; no request is now formally being made for phase three support.)
Additional funding may also be requested in phase three for continuation of research and development in other areas of information retrieval beyond the reference-retrieval application.

Other Users of SPIRES

Two other projects are now planning extensive use of the SPIRES programs for their reference-retrieval applications. One is the U.S. Office of Education's ERIC (Educational Resources Information Center) Clearinghouse for Educational Media and Technology, which is located in the Institute for Communication Research at Stanford. The ERIC Clearinghouse plans to input 1000 ERIC references through the SPIRES system and to perform a complete cost analysis before their decision on ultimate use of the system. The cost analysis should prove useful for SPIRES in its plans to phase into a self-supporting operation. The expectation is that the ERIC Clearinghouse will continue to use the SPIRES programs but with their own data base, after the initial test period (their alternative, namely batch searching via a commercial system, would be taken only if operating costs were too high under SPIRES).

The other project is the Stanford Library Automation Project which is currently working on automation of the library acquisition process as the first phase of their automation plans. Plans for development of the acquisition system include use of SPIRES data base and index-building programs and the SPIRES retrieval programs as the nucleus of that system. They would require an extensive modification and update procedure for changing document information in their "in-process" file and extensive output procedures for preparing appropriate ordering, purchasing, and accounting documents. Their planned investment in update procedures would then be incorporated into the SPIRES programs themselves, rather than duplicating that programming, or not having as general a system available in SPIRES itself. The Library Automation Project staff and the SPIRES staff are working in close collaboration during this development.
Facilities

The computer facilities at the Stanford University Computation Center, both at the Campus Facility and at the Stanford Linear Accelerator Center Facility, are adequate for the conduct of this project. It is planned to continue use of the 360/67 at the Campus Facility as the major computer for this project. The system software that makes such an on-line reference-retrieval application possible in the absence of a general purpose time-sharing system for the large machines of the 360 generation is the Stanford Terminal Processor used by Stanford's WYLBUR system of interactive text-editing and remote batching.

The facilities of the Institute for Communication Research are adequate for the conduct of this project. The project is housed, with the rest of the Institute (including the ERIC Clearinghouse) in a building near the Stanford Computation Center (Campus Facility). Previous experience of the Institute faculty and staff, who will be available for consultation, should be valuable. The Director of the Institute, Dr. Wilbur Schramm, is currently a member of the President's National Library Commission. Dr. William Paisley is Associate Director of the ERIC Clearinghouse, which is concerned with the flow of scientific and technical information among the educational media research community. Previous and concurrent Institute research on the needs and uses of scientific information and on the networks of communication among scientists provides an appropriate context for this project. Dr. William Miller, Professor of Computer Science and head of the Stanford Linear Accelerator Center Computer Group, will remain as Principal Consultant to the project. The two senior systems programmers now on the project staff provide the core of the necessary computer group. Louise Addis, a librarian at the SLAC Library, will continue one-quarter time with SPIRES, supervising the SLAC preprint collection input and SLAC Library liaison. One additional systems programmer will be hired, in addition to the hiring of graduate students. Graduate students in both the Communication and Computer Science Departments will be available to serve as research assistants.
Biographical Information

Edwin B. Parker, 36, is a naturalized U.S. citizen, born in Canada. He received a B.A. in Philosophy from the University of British Columbia (Vancouver, Canada) in 1954, an M.A. in Communication from Stanford University in 1958 and a Ph.D. in Communication from Stanford in 1960. From 1960 to 1962 he was Assistant Professor of Communication at the University of Illinois. He has been on the faculty of the Communication Department at Stanford since 1962 and is presently Associate Professor of Communication. He is co-author of *Television in the Lives of our Children* (Stanford University Press, 1961) and is co-editor of *The Kennedy Assassination and the American Public: Social Communication in Crisis* (Stanford University Press, 1965). He has published articles in *Public Opinion Quarterly*, *Journalism Quarterly*, *AudioVisual Communication Review*, *American Documentation*, *American Psychologist*, *American Political Science Review*, and other journals. During 1968-69 he will be a Fellow at the Center for Advanced Study in the Behavioral Sciences. (During that time he will continue active direction of this project.) He is a member of the American Psychological Association, the American Sociological Association, the American Statistical Association, the American Association for Public Opinion Research, the Association for Computing Machinery, and the American Society for Information Science.

William F. Miller, 42, received a B.S. in 1949, an M.S. in physics in 1951, and a Ph.D. in physics and mathematics in 1956, all from Purdue University. He was at the Argonne National Laboratory as a physicist from 1955 to 1959 and as Director of the Applied Mathematics Division from 1959 to 1964. During part of that time he was Visiting Professor of Mathematics at Purdue and Professorial Lecturer in applied mathematics at the University of Chicago. He has been Professor of Computer Science and Group Leader of the Stanford Linear Accelerator Center Computation Group since 1965. He has published articles in *Physical Review*, *Radiation Instruments*, *Nuclear Physics*, *Nucleonics*, and *Nuclear Instruments and Methods*, among others. He is a member of the American Mathematical Society, the American Physical
Society, the Association for Computing Machinery, and the Society for Industrial and Applied Mathematics. He has been a member of the U.S. Atomic Energy Commission Mathematics and Computer Science Research Advisory Committee since 1962.

Rutherford D. Rogers, 52, received a B.A. from the State College of Iowa in 1936, an M.A. from Columbia University in 1937, and a B.S. from Columbia in 1938. After military service and a period as research analyst on Wall Street, he was Director of the Grosvenor Library, Buffalo, New York, from 1948 to 1953. From 1952 to 1953, he was Director, Rochester Public Library, and Director, Monroe County Library System. He was with the New York Public Library from 1954 to 1957, first as Chief of the Personnel Office, and then as Chief of the Reference Department. From 1957 to 1964, he was Deputy Librarian of Congress. Since 1964, he has been Director of University Libraries at Stanford. He is (1967-68) President of the Association of Research Libraries. He is Chairman of the Association of Research Libraries Liaison Committee with the National Advisory Commission on Libraries. He has been active on the executive of the American Library Association, and as a consultant or member of numerous national and international committees or organizations dealing with library problems. He has published in the American Library Association Bulletin, College and Research Libraries, and Library Trends, among other journals. He is Principal Investigator for the Stanford Library Automation Project.

Allen B. Veaner, 38, received a B.A. from Gettysburg College in 1949, a B.H.L. degree in 1952 and a Rabbi degree in 1954, both from Hebrew Union College. In 1960 he received an M.L.S. degree from Simmons College. He was a Cataloger for the Harvard College Library from 1957 to 1959 and a Specialist for Document Reproduction, Harvard University Library, from 1959 to 1964. He was Chief Librarian, Acquisition Division, and Chairman of the Library Automation Committee at Stanford from 1964 to 1967. Since 1967, he has been Project Director for the Stanford Library Automation Project. He has been
active in the American Library Association and the National Microfilm Association. He was Chairman of the recent Library of Congress ad hoc committee to evaluate the proposed Library of Congress MARC tape format and character set specifications. He has published articles in Library Resources and Technical Services, in the Proceedings of the National Microfilm Association, and in the Journal of Medical Education.
March 7, 1968

Addendum to SPIRES Proposal

This document is submitted as an addendum to NSF development proposal number N-3042, titled "SPIRES (Stanford Physics Information Retrieval System)". It is intended to specify the technical plans for the first 12 months of development. Funds are requested for 12 months of development with expectation that funds for implementation of an operational system will be requested for an additional 18 months.

There are two main parts to the planned effort in the 12 months. One is the expansion of the present demonstration SPIRES into a useful pilot system on the 360/67. The other is the detailed system design for the larger scale operating information retrieval system, including hardware and operating system decisions, and decisions concerning expansion of the user population.

Preceding the pilot SPIRES is the continuation of development of the demonstration system described in the SPIRES 1967 annual report and demonstrated to NSF visitors. That system will serve as the nucleus of the pilot SPIRES to be developed and tested during the 12 month period starting June 1968. The overview of the system development plan showing target dates is presented in Figure 1.

Pilot SPIRES

The pilot system is planned as an interactive multiple-terminal reference retrieval system for physicists (primarily high-energy physicists). In addition to the physics data base there would be a generalized file capability to permit users from any discipline to input and retrieve document references in any subject they wished.

This combination of a large data base in a single subject area (initially the Stanford Linear Accelerator Center preprint collection and Nuclear Science Abstracts) plus smaller files in a wide variety of areas should provide the experience needed in the design of the more general operational system. The document reference collections in geology and in educational media research that were
Figure 1
Overview of System Development Plan

- **Data Base:** Input Processing, File Organization
  - Sept 67

- **Index File Organization, Processing**
  - Sept 67

- **Retrieval Programs**
  - Sept 67

- **Demonstration SPIRES (Batch)**
  - Oct 67

- **Demonstration SPIRES (2741 Remote Batch)**
  - Dec 67

- **Demonstration SPIRES Multiple Files**
  - Feb 68

- **Demonstration SPIRES Large Files, Tape Input**
  - May 68

- **PILOT SPIRES Multiple 2741 Terminals Real-time Interaction**
  - Oct 68

- **Specifications for Redesign of SPIRES**
  - June 69

- **Operational SPIRES (Initial Version)**
  - Dec 69

- **Demonstration SPIRES 2250 CRT**
  - Dec 67

- **Demonstration SPIRES Multiple User-initiated File Capability**
  - July 68

- **Decisions: Data Base Expansion**
  - Dec 68

- **Hardware and Operating System Implementation**
  - June 69

- **Hardware and Operating System Decisions**
  - Oct 68
demonstrated to NSF visitors were very instructive to the project staff. Minor problems were encountered, and corrected, in the different data bases that would probably not have been detected using only the SLAC preprint collection.

Nine tasks need to be completed to turn the demonstration system into a realistic pilot system:

1. Completion of disk input/output routines to permit expansion to large data base and index files;
2. Programming to overlay the retrieval programs to fit into a dedicated partition in the 360/67;
3. Programming for interactive (on-line) multiple terminal handling within the dedicated partition, interfacing with the typewriter-terminal driver programs used in the remote-batch demonstration system;
4. Completion of programming for updating records (correction, addition, deletion of bibliographic records);
5. Creation of a general multiple file capability for users who wish to input their own collection of references;
6. Programming to collect statistics on system efficiency and user habits and responses;
7. Inclusion of additional data bases to expand the physics reference retrieval capability;
8. Programming for additional indexes to improve the retrieval capability (e.g., citation indexes);
9. Programming to permit retrieval using IBM 2260 CRT display terminals in a multiple terminal environment.

Programming for the pilot system should be completed within the first six months of the proposed additional funding (i.e., by December 1968). Additional debugging, program checkout, test utilization, and analysis of statistics on system performance should continue during the following six months, while the detailed specifications needed to modify Pilot SPIRES into Operational SPIRES are being written.

System Planning

The major technical choice to be faced is the determination of the hardware and associated system software for the Operational SPIRES
to be created. The various alternatives can be subsumed under three categories: (1) to operate in a partition of the 360/67 (or its successor in the role of the main computer of the Stanford Computation Center's Campus Facility); (2) to operate in a second computer linked to the 360/67 (or its successor), possibly sharing peripheral equipment and terminal handling software; or (3) to operate in a separate computer dedicated to the information retrieval application. Several requirements will have to be kept in mind in such a choice:

(1) The operating information retrieval system must be economically viable.

(2) There must be reasonable guarantees of long term (i.e., at least five years or so) system stability.

(3) The system must be accessible from most (preferably all) remote terminals on campus, many of which will have been installed for other purposes (e.g., remote computing).

(4) Response times must be sufficiently short that users are reasonably satisfied with the services available. (This is, in part, subsumed under requirement one, in that a population of satisfied users is necessary to maintain economic viability.)

(5) There must be flexibility for continued expansion of the size and variety of data bases and the corresponding user populations from a variety of disciplines.

The choice should be made early in the development project, not later than the end of the first four months of development (i.e., October 1968).

**User Population**

The planned expansion of the data base of document references (and, consequently, of the user populations to be served) is indicated in Figure 2.

The first major group of users, as the project name implies, is comprised of physicists of which there are approximately 150 at Stanford. Physicists were chosen because of the relatively large number of potential users concentrated in a single discipline.
Figure 2

Data Base Expansion

SLAC Preprints (Small File)
Oct 67

SLAC Preprints (Large 2741 Input)
May 68

Multiple User-Initiated Files (2741 Input)
July 68

Other Users, SPIRES Demonstration
Feb 68

NSA Tapes (Large File)
May 68

Other Physics Tapes
July 68

Science Citation Index
Sept 68

Decisions: Choice & Timing of Additional Data Bases
Dec 68
In order to provide a general capability for users willing to input their own collection of references, a generalized multiple file capability will be provided. The variety of disciplines that are expected to make use of this capability should provide useful experience that will aid later system development.

The first major expansion planned for areas outside of physics is the Science Citation Index (available in weekly magnetic tape service) covering some 1600 scientific journals from a variety of scientific disciplines ranging from psychology to physics. This should provide a general journal article coverage of use to a larger population beyond physics. For example, some faculty members in the School of Medicine have expressed interest in having such a data base available.

Further expansion to services for other user populations (and, consequently, provision of other data bases) will require careful planning. As services are provided for each new user population, the information needs of those users will be studied to ensure that their needs for bibliographic information are being adequately met. This behavioral science research is considered necessary to provide criteria for evaluation of the system and feedback needed for efficient operation. The choice of which data bases should be added in which order will depend on two factors. One is the availability of machine readable data bases (e.g., Chemical Abstracts, Medlars). The other is the size of local user populations and their demand for such services.

In order to assist the project staff in determining priorities for development, a faculty advisory committee is to be established. The present intention is to expand the present Project BALLOTS (Stanford's Library automation project) faculty advisory committee to make it a joint advisory committee to the two projects. A plan for systematic expansion of services to the various user communities should be developed and approved by the advisory committee by the end of the first six months of the development project, i.e., by December 1968.
The budget for the 12 months starting June 1, 1968, is as proposed for the first year in the proposal to which this document is an addendum and is listed in detail in that document. Revised budgets for the effort after the first 12 months will be submitted at the appropriate time.