AD

OAR 69-0016
20 JULY 1969

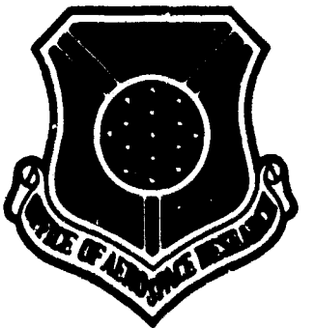
SELECTIVE DISSEMINATION OF INFORMATION (SDI)

Analysis of Experimental and Operational SDI Services, 1967

A. G. Hoshovsky

Published by

OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE



71

CONTENTS

ANALYSIS OF EXPERIMENTAL AND OPERATIONAL SELECTIVE DISSEMINATION OF INFORMATION (SDI) SERVICES

PART I - AN OVERVIEW

A. OBJECTIVES, SCOPE, AND METHODOLOGY

The impetus for the nation-wide study of SDI (Selective Dissemination of Information) came from our desire to learn more about the options available for the application of the SDI concept prior to our own decisions about its possible use in Air Force research and development activities. To this end we sought data on user populations, literature coverage, methods of establishing and maintaining profiles, as well as information on the supporting equipment and special techniques.

The findings and conclusions of this report have been drawn from the analysis of operational characteristics in thirty-eight (38) SDI services of the United States, and three (3) SDI services of the United Kingdom. This represented a major portion of some fifty (50) systems known to be in operation as of August 1967, and therefore can be taken as a fair sample of the total population.

Our principal data collection and verification tools consisted of structured questionnaires (shown in Part III), some interviews, and published literature on specific SDI systems (Part III). To insure accuracy of data we asked all of our respondents to review and comment on our first draft report. Their responses and critique have been incorporated in this report.

B. SDI SERVICE DEFINED

The first difficulty we encountered in the course of this study was defining the concept of SDI. The original concept was proposed, elaborated and demonstrated by the late H. P. Luhn in 1958. Luhn's notion was that computers were ideally suited for screening large numbers of indexed documents and matching them with terms which describe the scientist's profile. Such screening would be automatic and periodic; the scientists would receive "tailor-made" announcements about items of individual interest no matter where published and thereby manage to keep up with the literature despite the proliferation of journals and technical reports and at the same time save time and effort.

The idea of selective dissemination itself was not new, however, since it had already been widely practiced by many librarians, who could now call their service non-computerized SDI. Thus eventually the SDI concept came to embrace many non-automated systems which furnished the user with notices about new literature. It was the selectivity that attracted our attention. We therefore defined SDI as follows:

> SDI is a document-alerting service which selectively notifies users about new, or newly accessioned, literature. The selectivity is provided by matching subject areas of each document with user "profiles" which uniquely define users' areas of interest. Such a service has firmly established rules of document selection, profile matching, and user notification. The users serviced by an SDI service may be individuals or groups of individuals who combine their interests into a single profile.

C. MAJOR FINDINGS AND CONCLUSIONS

1. Diversity of Objectives, Outputs, and Procedures

If any generalization about SDI can be safely supported, it is that the surveyed SDI systems vary greatly in their design and operation. While meaningful practices of the majority can often be found, the exceptions are too significant to be overlooked. In fact these exceptions often point out very useful and unique applications of SDI systems. Most of the surveyed systems seem to be unique in at least one area, and the majority of systems varied from parameter to parameter. Some examples of the diversity among systems follow.

The existing SDI systems serve a great variety of users as well as a great number of users: scientists, engineers, managers, salesmen, doctors, and clients with any information need. Fully operational systems serve from 12 to 2800 users. The "users" are usually individuals, groups artificially created by the system, research laboratories, individuals forming their own group, administrative groups, and even companies. Altogether the surveyed SDI systems process anywhere between 25 to 30,000 documents per month. The SDI systems are performed as often as daily and as seldom as once every 2 months. Some SDI systems are automated while others are semi-automated or manual. While some SDI services depend exclusively on the bibliograhic tapes produced by major U.S. document centers, there are just as many who handle their own document accessioning and indexing. There is a great variety in construction of profiles. While some offer the user a choice of only a few subject categories, others use profiles of over 100 terms. Some use controlled vocabularies for both profile and document indexing, while others depend on free language indexing. Many profiles are being matched against processed literature by people while many others use elaborate computerized matching logic.

All this diversity demonstrates that SDI as a concept is easily adaptable to a great variety of objectives and operating conditions. Its potential applications have hardly been exhausted. For instance, one drug company uses SDI to keep its doctors and salesmen informed of company and competitors' products. The SDI concept has also been successfully applied to primary distribution of laboratory reports and is likely to effect the future pattern of technical journal production and distribution.

2. Similarities in Development Histories

With all this diversity of design there is a remarkable similarity in the origin of these systems. All but two have been established as the subsystems of other documentation services seeking to capitalize as much as possible on the existing equipment and procedures. Moreover, most of them have been developed as the services to the R&D community in Government and Industry. Finally, most of them began as "experimental" systems.

3. The Period of SDI "Adolescence"

Practically all operational systems started as experiments in selective dissemination so as to be better able to cope with the uncertainties of the new operational mode. We also suspect that in many cases the term "experimental" was really an euphemism for managements' unwillingness to face the consequences of an unsuccessful system. Since we have no data on the death rate of the experimental systems, we are unable to estimate the chances of any new system to reach the operational status. The available data does suggest, however, that most of the systems which began experimentally are still alive - mostly because they were subsidized by several federal agencies and major industrial organizations. There are, of course, notable exceptions: operational systems that are being supported through commerical subscriptions like the ASCA service of the Institute for Scientific Information and Scientific Documentation Centre, Ltd., in the U.K.

both have, however, a very extensive data base and hence offer a very broad field of vision to the potential subscribers.

Our data shows that on the average it has taken two to three years for the system's operators to gain sufficient confidence before they were willing to declare their system as operational. Yet there is also evidence of new semi-operational systems being established by utilizing many of the federally developed bibliographic tapes and the necessary SDI computer programs. This, coupled with the evidence of a reasonably rapid growth of the new experimental systems, a largely untapped market potential and increasing availability of experienced SDI managers, leads us to believe that the "experimental" period in future systems may be considerably shortened. However, it is expected that several years will lapse before it is possible to see a marked decrease in time of experimental systems because of the unique requirements at each installation. These special requirements will probably continue despite access to large data bases (Defense Documentation Center, Chemical Abstracts Service) and software packages.

4. Applications of SDI Techniques

The early SDI concept emphasized benefits to be derived from SDI as a means to scan automatically the whole of literature in search for items of user's interest. By doing so it would provide the user with the ability of learning about pertinent papers published on and beyond the "peripheries" of his regularly-used literature.

All of the surveyed systems provide the user with this vision - but with some important differences. At least five variants have been noted and they deserve special mention. To simplify our descriptions we have named them:

a. Peripheral Vision SDI

b. Focused Vision SDI

c. Dissemination Aid SDI

d. Library Acquisitions SDI

e. Group Profile SDI

The Peripheral Vision SDI needs little comment since it is essentially an SDI service in which the whole universe of available documents is being scanned and provided to the user according to the predetermined hit rules. SDI systems using NASA or Chemical Abstract tapes are prime exponents of this type of service.

The Focused Vision SDI is a variant which is practiced in a varying degree where the literature input is being evaluated not only as to its relevance to the general area, but also for its scientific or technical worth. In such a system the user is relieved of the need to decide which of the papers are of sufficient value to be ordered - the systems operators act as his filters for much of the material which otherwise might have resulted in a hit. Because of this prescreening the user usually receives only notices on those items which are believed to be most relevant to his needs - his attention is being "focused" on a much narrower segment of the existing literature which may lie on the peripheries of his usual literature sources. The Bureau of Reclamation for instance, does such a thorough job of prescreening documents that only 5% of received documents is entered into its data base.

Dissemination Aid SDI has been observed as an attempt at providing a more accurate tool for the dissemination of the internally produced literature. When it is so used, the SDI does indeed provide a kind of peripheral vision to the user since it can be easily argued that the company-produced literature usually lies outside of his regularly scanned sources. In one innovative approach an aerospace company is using the SDI method to its commercial advantage by insuring that the "right kind" of technical literature is sent to "appropriate" Government

R&D managers. By holding up before these managers the relevant company work it is hoped to enhance the company's image in a given technical area and, consequently, increase a probability of future R&D contracts.

Library Acquisitions SDI. This variant was found in only one of the surveyed systems but it is a most interesting form of SDI. Basically it serves both as a screening device for the research laboratory's library acquisitions as well as a method of notifying the library users about the newly received documents. This variant employs profiles of its users to determine the over-all library profile against which is being matched bibliographies on magnetic tapes of major U.S. documentation centers. Of equal, if not greater interest, is the fact that this is the only known system which has justified its worth by contributing to the reduction of technical library acquisition costs. By avoiding purchase of marginal-interest literature, the cost of SDI is being rationalized by showing the evidence of savings which are possible by not buying and storing "useless" documents.

Group Profile SDI. SDI was originally created to provide service to individuals, by keeping them up to date on secondary fields of interests as well as in their own occupational fields. The individual interest profiles tend to be costly however, and require a lot of planning and a large data base. Some organizations unwilling to spend the time and money necessary for such personalized services have moved in the direction of group profiles where each profile represents the combined interests of a single professional group. The groupings can follow several distinct lines: groups involved by a single research project, groups having common academic interests, or groups where the members occupy similar positions in the company, etc. (like salesmen of a given product line).

5. Most Services are Under-Used

There are two significant parameters in measuring the use of a system: system capacity (limits determined by configurations of a system) and the number of eligible and available users that participate. The surprising findings of this survey are that all systems are not even close to their capacity. Moreover only a relatively small portion (22%) of the eligible users depends on the SDI services. Despite this apparently weak showing, all systems surveyed expected to continue their service. It should also be noted that four of the big complex systems operate with between 80% - 100% of their eligible users.

Why the many services are not more fully utilized cannot be answered with our data. At this time we can only postulate that the inadequate advertising of the services is one of the contributory causes for the minimum use of the SDI services.

6. Optimum Sizes of User Population

In principle SDI concepts and methodologies can be made to work for any number of users. The wide range number of users found by this survey (15 to 2800) tends to support the validity of this contention.

But let's be more specific about the optimum number of SDI users per service. Our study shows that experimental and partially operational systems usually work with a small number of users. The average found in this survey is 32. These users supply a considerable amount of feedback during this trial period until the system is ready to run. Fully operational in-house systems usually operate with 400 or fewer users.

The reasons that systems have less than 400 users seem to come from other than SDI limitations. Most SDI managers feel, for instance, that system capacity was not being approached and their systems did not require changes to accommodate a marked increase in the number of new users. Reasons that are more plausible are likely to be machine and staff time, priorities, the kind of subject matter selected for document inputs, and perhaps

the interest of eligible users. Our survey indicates that a surprisingly small percentage of users (less than 25% of eligible users) takes advantage of the SDI systems. Two systems indicated that concerted advertising was necessary to get people to become active users. In one system only one-third of the eligible users were initially interested, but after the rest saw the results, the number of users rapidly increased to three-fourths.

7. Sources of Input

It is possible to operate an SDI system without having an in-house library or technical information center. About two-thirds of the surveyed systems use other sources for at least part of their document input. Half of these systems are dependent on these sources for 100% of their document input. Two options were indicated by participants: dissemination centers that usually supply magnetic tape, and contractors that perform indexing and other tasks to prepare machine-ready records or that perform the entire SDI function.

Using other sources for document input has several significant advantages. Selecting and preparing documents for SDI is the biggest on-going task in this process. The time, personnel, and equipment savings are immediately obvious. In addition using a dissemination center's input will in all probability put some constraints on the system options and thereby cut down on expensive experimentation. Dissemination centers often offer software packages and consulting services, or they can perform the entire service in-house. Using a contractor offers the same time, personnel, and equipment savings but at the same time allows installation control over the SDI system development.

One system saves indexing personnel by using only author-indexed papers. The system, which uses only in-house [illegible], inputs the author's index terms and delivers to the author a list of names to whom his paper should be sent. The author is responsible for sending his paper to those people.

8. Controlled vs. Free Vocabularies

A significant majority of SDI profiles is being constructed with the aid of some form of controlled vocabulary. This majority (over 80%) is in turn divided into two-thirds using its own thesauri, and one-third using a published version by such organizations as DDC, National Library of Medicine or the Engineers Joint Council.

Of the computerized systems only four use free language input. Document input for such systems is simple and allows an installation to input documents from several sources without restructuring data. User profiles, on the other hand, become long and complicated. The systems work efficiently, but systems using controlled vocabularies are easier to operate.

In little more than one-half of all systems the users are not restricted to any limit of the technical terms they can use in their profile. In practically all cases profiles are edited or reworked by the systems operators into controlled vocabulary terms.

Only one system practices "autoprofiling", where the profile is constructed without the direct choice of the profiled people. The autoprofiling concept entertained by various people in the past is apparently not finding wide acceptance.

9. Individual vs. Group Profiles

All surveyed systems managers believe that SDI serves a real and important need of screening the literature for specialized users, and they intend to continue some form of current awareness.

There are some signs that economic

factors attract certain SDI operators to move more in the direction of group profiles and away from the personalized service where each collection of documents is uniquely matched with the individual's profile. At least one large system (NASA) is having second thoughts about the economic feasibility of a large, fully individualized SDI service and has developed a new program called SCAN in which computers are used to prepare specialized, periodic announcements based on group profiles. Similarly, our own office stimulated a test of the group profile approach which resulted in the development of the Clearinghouse notification service known as CAST (Current Announcements in Science and Technology). There is also evidence that many individuals receiving personalized SDI services do in fact represent group interests. By establishing their profiles so that they encompass the interests of a closely related group, they act as the front men for the service for a number of individuals. This last phenomenon is looked upon with favor by many systems managers who see in it the means of reducing machine processing and mailing costs, and, hence, user costs.

Despite the seeming attractiveness of group profiles only 25% of the systems use them; the remaining 75% use individual profiles as the base for their services. It must be remembered, however, that most of the presently operating systems have no economic incentive for any departure from the individualized profiles. It is interesting to note, however, that in several cases where the user was given a choice of more expensive individualized service and a considerably cheaper group profile service, the switch to the cheaper method runs as high as 70%.

10. Operating Costs

As might be expected there is a tendency for yearly costs per user to decrease as the number of users increase, although not at an even rate. Of the eleven systems that supplied accurate annual budgets, the average cost of five systems with user population of 200 or less was $234 per user. The average cost of systems with population of 200-2000 was $115. Although this tendency is worth noting, the sample in our survey is too small to place too great a reliance on the cited costs.

Since SDI services are usually provided as an adjunct to other information services, and since we were unable to find a typical system, the survey did not permit an estimate of the typical costs required to establish a new SDI service. The analysis of budgets (where available) and its correlation with other data permitted us, however, to construct an index which could prove useful to future systems planners. This index is derived from the calculation of the average cost per user per run, which in our survey is calculated to vary between $1 and $2. This index, unfortunately, carries some hidden relationships that could not be broken out for analysis. The number of profiles and documents, matching techniques, and systems maintenance contribute to the cost, but none appears in the index. Perhaps the best way to use this index is to determine the number of users and the SDI budget, then to devise an SDI service and determine its cost per run. The index can be used to calculate systems costs likely to be incurred in the operating year.

11. Profile Matching Techniques

Matching is the focal point in SDI. At this point the user profiles and documents are matched to determine what documents are relevant to which users. There are four major matching techniques: individual or aggregated profiles matched against either individual or aggregated document profiles. (Aggregated means a master list of index terms, usually with appropriate tags to get back to the proper user or document.)

Three major matching strategies are

being practiced: linear, boolean logic, and weighted terms. The purpose of matching is to find "hits", i.e. events where a document term satisfies the user's profile term. Linear matching means that if any term in the user profile matches any term in the document record, then a hit is generated. Boolean strategy uses statements of terms connected by and, or, and not. A hit requires that the entire statement be satisfied. Weighting indicates that the terms of the user's profile are weighted to indicate relative importance. Matching operates in the same manner as linear matching except that a running score of weights is kept. The number of hits required to generate a notice is arbitrarily determined by each system.

Approximately 75% of the systems matched individual profiles against individual documents. This approach is advisable for any matching strategy other than linear matching, and most of the systems with boolean logic or weights use it. Linear matching can be performed almost in any manner, but individual profiles against individual documents is preferred by the surveyed systems. To use aggregated lists simplifies the matching step but adds additional sorting runs to find users and remove duplicates.

12. User Notices

The typical user notice contains an abstract and is machine-printed. Half of the systems use punch cards or double printed punch cards. Some of the remaining systems use cards of other than punch card size. Cards are probably preferable from the user standpoint because they are more durable and easier to file.

In general, users prefer an abstract on a notice because the user would rather make the final decision about the document's relevance. In addition since the document should be of at least some interest, reading the abstract may often supply enough information without taking much of the user's time. Therefore, document delivery as the form of the notice does not seem to be desirable from the user's viewpoint. Systems managers are probably just as happy because document reproduction costs can mount up quickly for a system with more than a small number of users.

Three systems deliver documents as the user notice. These systems are designed in such a way that document delivery seems to be the best approach. One system uses SDI to route selectively technical in-house papers. Another system uses task profiles to screen incoming documents and routes 25% directly to the tasks. The third system uses an SDI approach to route engineering drawings.

13. Frequency of Notification

Most fully operational systems distribute documents either weekly or every two weeks. Weekly notices are predominant. A few send daily notices, and a few send monthly notices.

Frequency seems to depend on system design and the user's expressed needs rather than document volume or SDI budget. SDI system design no doubt reflects the time made available for the service as well as the interpretation of the users' information needs.

Experimental and partially operational systems operate in a variety of ways. Some send very frequent notices as different parameters are tried on the same data base. Some systems distribute notices much less frequently than they intend to when the system reaches operational status.

14. Profile Modification

All SDI systems have provisions for profile modification. All surveyed systems not only modify profiles on user request, but also have some routine method of modification which does not require direct user objection to non-rele-

vant announcements.

The majority of systems (80%) relies on users to indicate profile changes. Most of these systems ask users to mark the returnable portion of the announcement. Approximately 20% of the systems initiate profile modification by scheduling modification at fixed time intervals.

Only one system has an automatic modification scheme. Whereas users in other systems indicate the changes to be made, users of this system indicate how appropriate the announced document is. The system modifies weights of terms relevant to the notice and terms that matched but did not generate a notice with no further interaction with the user.

D. SELECTED EXAMPLES

In this section we present brief descriptions of several operating systems which could be considered as useful prototypes for future SDI systems.

1. Computerized Focused Vision SDI

The system at the U.S. Bureau of Reclamation has about 2000 junion and senior civil and mechanical engineers and physical scientists. Each month about 1250 documents are reviewed by scientists in the field, and about 50 (4%) are inputted into the SDI system. Inputting averages 6 1/2 hours per document. Document indexing averages 20 terms.

This system inputs only 4% of the documents reviewed so that only particularly cogent documents will be inputted. Men who are professionals in their field read and select appropriate documents, and these documents are passed on to superiors who also eliminate documents. Hence, the documents finally inputted are highly specialized and "focused." This closely controlled type of input is apt to result in a high input time.

User profiles are constructed using the Bureau of Reclamation Thesaurus. The profiles are limited to 20 terms and no more than 4 of these terms may be coded with an * to indicate special importance. The user constructs his profile, but it is reviewed by the staff and altered if necessary to reflect what they believe his needs to be.

Matching is linear. If 3 terms or 1 special term match the document terms, a notice is generated. SDI runs are made monthly. The user receives an abstract or a punch card and a form for requesting the full document. The system provides 100% document and microfiche back-up.

Profile modification is primarily handled by staff who analyze the results of each output. User complaints are treated, but they are usually rare.

2. Manual Peripheral Vision SDI

The Scientific Documentation Centre based in Dunfermline, Fife, England, is a commercial SDI system. The system has a particularly large document volume, 300,000 items yearly, selected from a wide range of sources and countries covering a wide range of topics. Documents include all of U.S. Government Research and Development Reports and appropriate parts of U.S. and British Thesis Titles. The system started four years ago solely for SDI. Gradually the topics are being made available for retrospective searches also so that about half are included in a storage bank at present. Science, technology, medicine, botany, zoology, and engineering are some of the fields included. Fees vary according to topics selected. Most topics cost between fifteen and fifty pounds per year. Many of these prices are less than American systems.

A potential user reviews the list of topics available and indicates his choices. The staff at the Centre constructs the profile. The user receives

weekly notices on a 5x5 card or a punch card according to his preference. Each contains a citation and keywords. The user may order the document, but ordering documents requires an additional fee.

Specific facts about users and profile construction are considered confidential, as are the workings of the system. This system is operated without benefit of computers. A staff of forty, mostly students who work part-time, maintains the SDI service, establishing that manual systems are not limited in size or capability.

5. Group Profiles

NASA/SCAN is an experimental system that permits selective dissemination to large numbers of users in an efficient and inexpensive manner by modifying some of the standard SDI techniques. Instead of creating individual profiles to meet the information needs of individual users, SCAN offers a list of 189 topics, the profiles for which are already constructed. A potential user selects his areas of interest from the list and informs NASA of his decision. This type of approach seems to satisfy both the users and the systems staff. Profile construction is quite often very time-consuming and expensive because, although the user knows what he wants to receive, the user cannot always reflect his needs through index terms he chooses. His profile begins to reflect his needs after several SDI runs and conferences.

The user selects one or several of the SCAN topics, and his name is added to the list for each of the topic profiles. He spends a minimal amount of effort and time to participate in SCAN. At present 500 users interests are reflected by 189 profiles. Profile modification merely involves adding or deleting topics and can be done at any time.

The data base is made from articles in STAR (Scientific and Technical Aerospace Abstracts) and contain about 5500 documents per month. These articles require 1.5 hours of input time per document which includes abstracting.

Document records and interest profiles are matched twice a month. The interest profiles use weighted terms to simulate boolean logic. A notice is generated when the document terms satisfy one logic statement in the interest profile. A user notice contains the document citation and index terms with a returnable document request form.

4. Library Acquisitions SDI

Fort Belvoir (U.S. Army Mobility Equipment Research and Development Center) has an approach to SDI unique to this survey. This system pays for itself by screening potential documents to include only relevant documents in the Technical Information Division library. The money saved by not purchasing irrelevant documents more than pays for the system.

There are approximately 225 research tasks at Fort Belvoir. Interest profiles are created to reflect the work of each task. The profiles are constructed using an internal vocabulary based on the DDC thesaurus. The profiles are cumulated into a master profile for the division, and all documents are matched against the profile.

Documents that are accepted into the library are re-evaluated by a staff of technical information specialists. These people decide whether to route the document to the appropriate tasks, announce it in a bulletin of recent acquisitions, or both. Documents (not notices) are routed directly to the tasks if they adequately meet the task requirements. Documents of a more general but related interest are announced in the bulletin.

5. Free Language Computerized SYSTEM

The SDI system at Ames Laboratory,

USAEC, Iowa State University, can be considered a functional experiment for a regional and/or national non-specialized current awareness service. Few limits are put on the user or on document input; cost is low (between 3¢ and 10¢ per notification), and participation is good. In 1965 participation jumped from 1/3 to 3/4 of eligible users (senior scientists only) within six months after the scientists saw the benefits to be gained. Recently Ames SDI coverage has expanded to include industrial departments and corporations (included by the State Technical Services Act) and university departments, as well as individual scientists. Currently there are 220 interest profiles.

Profile vocabulary is not controlled. Users may list authors, journals, foreign language terms, synonyms. They may use word clusters of up to six words. They may use truncating and extension. They may use negative words and word clusters used for total negation. Terms are weighted between 0 and 1 to four significant digits, initially by the user and subsequently modified by the system. Long profiles are encouraged, and the average is 130 terms (words or clusters).

Document input, 6500 - 7500 documents per week, comes from many sources; therefore input is not uniform, but ranges from title and author only to abstract to text. There is no practical system limit to the length of a document. The system has been designed to adapt to assorted input by varying the hit level document by document (different from any other system so far) according to the number of terms available for matching. If only titles are given, the level is 0.3000; for abstracts the level is 0.5000; and when texts are used, the hit level is 0.7500.

After the user returns his Port-a-Punch card indicating his interest in the cited document, the weights of his profile terms are increased or decreased or are left unchanged automatically by a computer program used for profile modification. For a positive response, the weights of the matched user profile terms are increased; for negative responses, the weight of the matched terms are decreased. When profile terms match document terms but do not generate a hit, the weights on these terms are increased but at a slower rate. This method of modifying weights is so sophisticated that those users who wish to adjust their own weights soon give up and let the system do it. The breadth of document coverage is too much for one man to try to analyze and control the weighting strategy of his profile.

Ames' SDI system takes 3 1/2 hours per week running time on an IBM 360/50 and IBM 1401.

PART II - STUDY DATA

A. DETAILED ANALYSES OF DATA

The SDI questionnaire found in Appendix A, has 8 sections that pertain to 8 parameters to be analyzed by this survey: general factors, users, documents, profiles, profile modification, matching, user notice, and equipment and personnel. This portion of the report follows the format of the questionnaire.

1. General Considerations

This section covers background information for each SDI installation.

Questions 1 and 3, name of organization and formal title of SDI system, were only for records and will not be discussed. The remaining questions will be presented as follows:

Question 2. Each organization's definition of the SDI concept.

Question 4. Length of time SDI systems have been in operation.

Question 5. SDI budget.

Question 6. SDI as part of library or larger information center.

Question 7. Operational level.

Question 2. Each organization's definition of the SDI concept.

It was expected that major differences between systems could be traced back to each organization's interpretation of the SDI concept; hence the definitions supplied were to be used as the framework for analyzing each system. However, in most cases formal generalized definitions rather than individual working definitions were supplied, and no real insight into the uniqueness of each system was gained. All definitions received did fall within the limits of the survey criteria. The individuality and purpose of each system were more apparent from answers to the rest of the questionnaire and through discussion with participants.

Question 4. Length of time SDI systems have been in operation.

Two formal systems included in the survey, Crerar Library (System 4) and M&T Chemicals (System 38) have been operational prior to the name SDI. Since 1958 when Luhn presented the SDI concept, the number of systems has increased steadily each year. Figure 1 shows the cumulative growth of SDI since 1958 as represented by thirty-seven of the surveyed SDI systems. The graph shows both the number of systems operating each year and the number of new systems added each year. One system, B.F. Goodrich (System 37), is excluded from the graph. The system has been fully debugged and is capable of being fully operational but at present is non-operational owing to lack of available personnel. U.S. Army Natick Laboratories SDI system (System 9) is non-operational at present but is nevertheless included in the growth curve. Natick Labs has just completed a nine-month pilot test started in the fall of 1966, and the results are being analyzed at present in order to implement a fully operational system.

At present we believe the number of SDI systems is small compared to the number of potential systems. The service

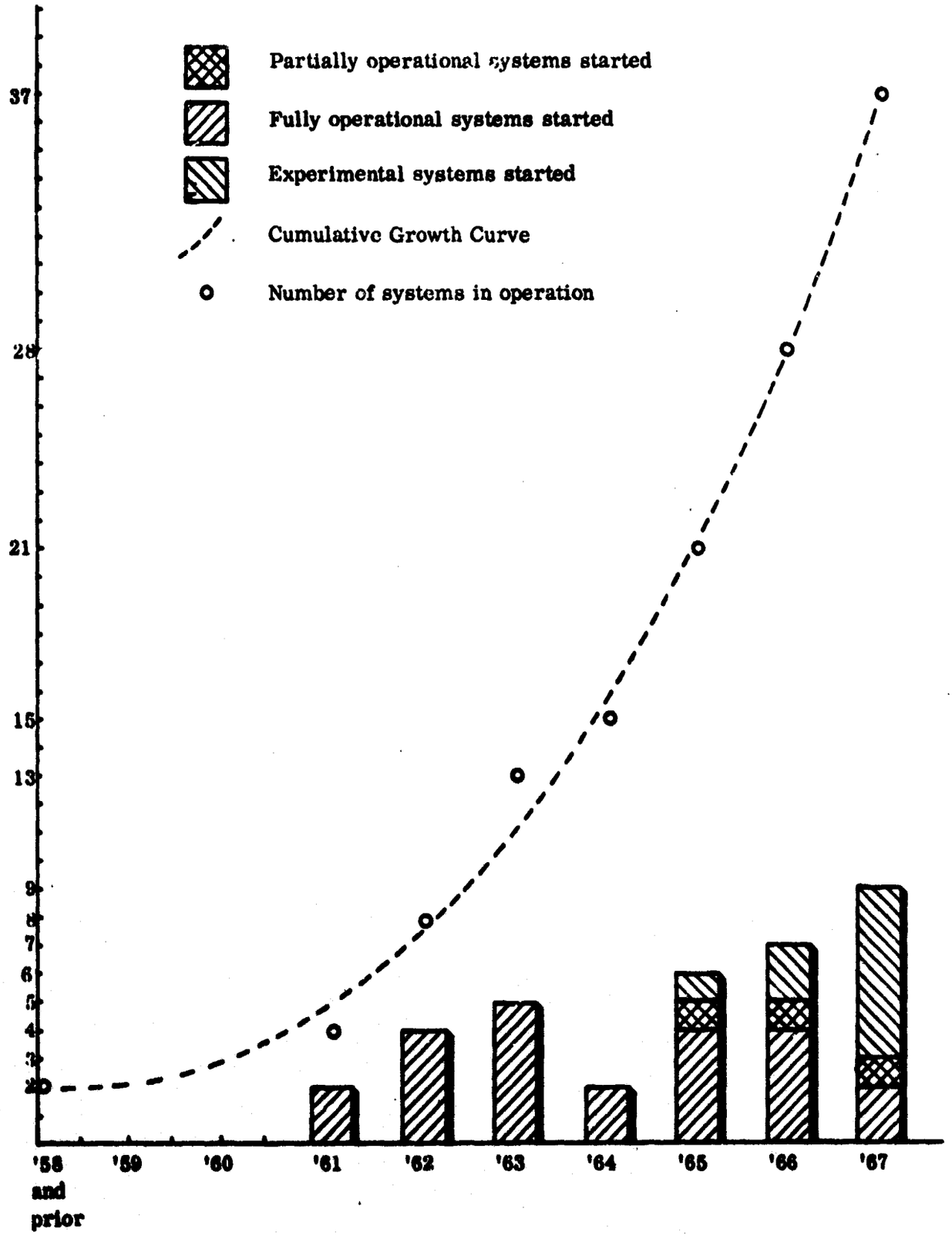


FIGURE 1. Growth Pattern of Surveyed Systems

provided by SDI seems to serve a very important need. We believe that SDI systems, or at least the service provided by SDI, will continue to be developed. Our opinion is supported by survey participants, some of whom expressed dissatisfaction with and expected to change operations, but not cancel the system.

Question 5. SDI budget.

Yearly budget figures were received from seventeen (45%) of the thirty-eight systems, with government systems supplying the majority. SDI installations were divided into three categories: government, private institution, and university SDI systems. (See Appendix B).

Budget figures were received from nine of the fifteen government systems, seven of the twenty-two private institutions and the one university system. The budgets range from "free plus two hours of clerical time per month" to $180,000 per year. The "free" SDI system belongs to the U.S. Naval Weapons Laboratory. For their willingness to take part in the NASA/SCAN experiment and supply NASA with statistics and other feedback, NASA is absorbing the cost. The only cost incurred by the Weapons Lab library is for one hour of clerical time for each SCAN run, made twice a month, to mail notices to users.

In some cases we were told that budget information was private and could not be released. In other cases SDI was an integral part of a larger system, and the cost could not be broken out of the total budget. Even for the budget figures received, compatibility between the figures was difficult to judge. An SDI system is often planned around existing equipment, programs, personnel, and time. Hence cost figures may include salaries in some cases, partial salaries in others, and no salaries in others, depending on how each installation interprets SDI cost. A detailed cost analysis of data received is presented in Part VII.

Question 6. Part of a larger system.

SDI systems are subsystems of a library or data storage bank. Two exceptions were found in this survey: National Cancer Institute and Vitro Laboratories. The former developed SDI as a primary system; the latter operates a manual SDI system primarily to distribute engineering drawings outside of its computerized information center.

Question 7. Operational level of SDI systems.

B.F. Goodrich's inactive but fully operational system is excluded from any of the three choices this question offers. The remaining systems are divided as follows:

Fully operational	65%	(24 systems)
Partially operational	8%	(3 systems)
Experimental	27%	(10 systems)
TOTAL	100%	37 systems

Comparing lengths of time experimental and partialy operational systems have been functioning seems to indicate that SDI systems need between two and three years to become fully operational. Since SDI is a new concept and the number of fully operational systems is relatively small, there is little precedent on which new systems may rely. Hence what seems to be a long period of trial and partial operation becomes a necessity, as most systems experiment with several SDI components before developing fully operational systems. One of the more interesting and significant findings in this survey is the variety of approaches to SDI that have been developed.

2. User

The following topics are covered:

1. Population limit of the system

and the present number of users

2. Population composition - Managerial and Other

3. Number of eligible users

4. Profiles reflect interests of

   a. Individuals and/or

   b. Groups related by

      1. Discipline

      2. Task

      3. Administration

Question 1. Population limit of the system and the present number of users.

Population limits

It is assumed that each SDI system has some finite user limit beyond which the system becomes impractical to operate in the present manner on existing equipment. This question is designed to evaluate system utilization to aid in the determination of the acceptance of SDI. The thirty-eight participating systems answered the population limit portion of the question as follows:

Thirty-two systems were open-ended.

Six systems had a fixed population limit.

The thirty-two systems have modified the meaning of open-ended. These systems are designed so that the present number of users and the anticipated number of users do not put any strain on the system. Hence it is not necessary for these systems to consider a maximum number of users. Two systems that did define system capacity, Bonneville Power Administration (System 17) and Ames Laboratory (System 20), should both be considered open-ended within the context used by the thirty-two systems above. Both operate significantly below capacity at 6.8% and 2.2% respectively; considering both systems operating with maximum eligible users, the percents of capacity are 30.0 and 5.0 respectively. These percentage figures are low enough to justify including these systems in the group with "open-ended" user limits.

Four of the six systems that defined population limits were not referring to system capacity. These four experimental systems have a fixed user limit only for the duration of pilot studies in order to control effectively a manageable number of users for test purposes. All four systems will become "open-ended" when they become fully operational.

Our analysis indicates that the SDI systems in operation expect to be able to serve many more users than they anticipate having. Part of the reason may be derived from the fact that SDI is usually a subsystem. The equipment used by an installation is large enough to accomodate the main system so that system capacity is much larger than SDI requires.

Population Size

SDI systems have been established to serve a variety of population sizes. The broad range of user populations indicates that SDI itself does not precipitate a model population size; rather any group of potential users may devise an SDI system to meet its needs. In order to analyze the number of SDI users meaningfully, the measurable SDI systems were divided into three groups.

Experimental and Partially Operational	13 systems
Group I, Fully Operational	16 systems
Group II, Fully Operational	7 systems
TOTAL	36 systems
No response	2 systems
	38 systems

Generally experimental systems operate with a small number of users. The users are specifically selected to supply statistics and other feedback to the system designer that will be unnecessary once the system is fully implemented. Ten of the thirteen systems have small user populations with a mean average of 32 users. The other three systems have populations of 500 (NASA/SCAN experiment), 300 (Naval Weapons Laboratory involved in NASA/SCAN) and 800 (Bell Telephone Laboratories' partially operational system).

The fully operational systems were ranked in order of user size excluding two systems that were unable to answer. There was a natural break between 400 users and 700 users. Group I includes 70% of the fully operational systems and has a range from 12 to 400 users. Group II contains seven large systems with populations widely scattered from 700 to 2800 users and are considered, therefore, atypical, but nevertheless significant.

The largest system in our survey, IBM's corporate-wide system located at Armonk, New York, serves as a good network for distributing IBM information to all interested personnel. It is likely that there are not many installations that need an alerting system this big. However, the IBM system exists and therefore establishes the feasibility of such a system.

One other system has been designed to establish the feasibility of large centralized dissemination centers, in this case unrestricted in subject coverage (Navy Laboratory with system capacity of 11,000 users). It may be that effective centralized dissemination systems will be established in the future, and with time sharing becoming more common perhaps replacing individual SDI installations. But for the present time at least our survey indicates that most SDI systems serve 400 or fewer users.

Cost and Users

There is a slight tendency revealed by our survey that as the number of users in each SDI system increases, the cost decreases. Fourteen systems could be tested, and we found that systems with 200 users or less have an average yearly cost of $234, and systems with 220 to 2000 users have an average cost of $115. We believe that this finding is somewhat inconclusive because the user is not a significant factor in SDI systems when considered separately.

Question 2. Population Composition

The purpose of this question was to ascertain what portion of SDI users were managers. The original choice offered to participants, Managerial or Bench Scientist, was too restrictive. Our survey found that SDI is not limited to scientific personnel in research laboratories but also includes doctors of medicine, engineers, and salesmen, as well as personnel in other fields where Bench Scientist does not accurately describe the user. As participants were interviewed, this question was presented as Managerial or Other, and the change was made on subsequently prepared questionnaires. In addition participants were told that the term Managerial was not limited to administrative personnel. For instance, it is assumed that research projects staffed by junior and senior scientists have a "Manager" who is himself a scientist.

A total of 82% of the surveyed systems serve both managerial and non-managerial personnel. Of the remaining seven systems two systems (Systems 3 and 4) distribute documents to managers only, and five (Systems 7,9,17,28 and 32) distribute to non-managerial personnel only. Both systems distributing to managers only use group profiles, and the managers redistribute notices within their respective groups. Of the five groups that distribute to non-managerial personnel, three will distribute to managers as well when the systems become

fully operational (Systems 9,28 and 32).

Evaluating all the responses both managerial and non-managerial personnel take or will take advantage of SDI services in thirty-six of the surveyed systems. Evidently there is enough of a common interest between these two groups that both take advantage of SDI, or else the scope of the data base is broad enough to meet the needs of all types of users.

Question 3. The number of eligible users.

It is desirable to compare the number of users to two ways: with respect to system capacity (Question 1) and with respect to eligible users. Eligibility is determined by each installation and is considered as separate from system capacity. Criteria include education (one system for instance, requires an MS or better), field of interest (limited by document input), occupation and job level. Some installations had no real requirements except that the users be professionals. Seven of the systems are designed in such a way that the number of eligible users is not relevant. They are systems that work for clients (e.g., Crerar and Aerospace Research Applications Center) or that are open to very large segments of the population (e.g., the SDI system at National Cancer Institute will be open to all biomedical scientists involved in cancer research).

The question was applicable to thirty-one systems, but five of these systems were unable to supply data. The percent of eligible users that use SDI was determined for the twenty-six measurable systems.

4 systems	80% - 100% of eligible users
22 systems	mean of 18%

It would seem that SDI should be a boon to persons overloaded with new literature and the need to keep informed. However, only four systems include between 80% and 100% of eligible users. The Bureau of Reclamation (System 18) includes 80% of their eligible users and is the one system of these four that has only individual profiles. The Bureau is also the largest system in this group with 2000 users and 2000 user profiles. Western Electric Company (System 24) includes 83% of their personnel, 2000 users. Because Western Electric uses group profiles however, the system has only 300 interest profiles. The remaining two systems, Fort Belvoir (System 35) and IBM at Poughkeepsie and Kingston (System 34), operate at the 100% level and use group profiles.

The remaining twenty-two systems establish that SDI is not widely accepted. We removed the experimental and partially operational systems from the group, and the mean percentage of eligible users rose from 18% to a little less than 22%. As stated previously, experimental and partially operational systems intentionally work with a small number of users.

We found that one possible reason for low participation rate is lack of exposure. When the SDI system at Ames was first proposed, little interest was expressed by eligible users and only one-third became users. After the scientists saw results from SDI, the number of users rapidly increased to 75% of eligible users. Those involved with the SDI system at Fort Monmouth developed a promotional campaign to generate enthusiasm and exposure. Perhaps Fort Monmouth's approach may well be advised for large organizations.

We also found some error in the figures supplied by at least three and perhaps as many as eight systems that belong to the low participation group. These systems use group profiles and call "users" those people to whom notices are sent, rather than all of the people in the groups who receive benefits from SDI. There is some reason to believe that

eligible user figures supplied by these systems reflect the number of people rather than the number of groups that are eligible for SDI. Hence, the figures cannot be compared because of the difference in units. The percentage of participation would definately be raised if these figures were corrected, but we cannot determine whether or not the increase would be significant.

Question 4. Group or User Profiles

SDI was originally designed as a service to individuals so that each profile was to reflect the information needs of an individual. However, profile construction is time consuming and expensive enough that some installations use profiles to reflect the needs of a group.

Many of the SDI systems (about 75%) use individual profiles either exclusively or for most of their profiles. When these systems use group profiles, it is by user's choice.

Group profiles, used by 25% of the surveyed systems, provide selective dissemination more economically. For instance Western Electric (System 24) can meet the needs of 2000 people with only 500 profiles. Profiling time and computer time are saved. Most systems using group profiles base profiles around task groups.

Group profiles are easier to handle from the systems staff viewpoint, and there is some evidence in our survey that users may prefer group profiles. One commercial system that had been operating only with individual profiles developed a package similar to NASA/SCAN and offered both to their customers. (See IV. Selected Examples for discussion of NASA/SCAN.) The system cost for group profiling was reduced, so the savings were passed to the users. There was approximately 70% changeover from individual to group profiles. The users did not mind losing options available with individual profiles as much as they saved money and profile construction time and then informed future systems planners to consider group profiles as a less expensive method of meeting user needs.

3. Documents

The questions found in this section of the SDI questionnaire will be discussed as follows:

Questions 1 and 10. Systems relying on dissemination centers for document input. Systems contracting for indexers.

Question 2. Average document input time.

Questions 3 and 4. Number of documents included as SDI input per month and what part these are of total documents reviewed.

Questions 5, 6 and 7. Documents indexing procedures: vocabulary and limit and average number of index terms per document.

Questions 8 and 9. Kind of documents used for SDI document input.

Question 11. General subject areas of input.

Questions 1 and 10. Systems relying on dissemination centers or contracted indexers for document input.

Selecting, indexing and abstracting documents are the biggest on-going tasks involved in dissemination of current documents. When document input can be acquired from another source, the system saves time and personnel. We call these sources dissemination centers in this report, although in some cases dissemination is not a main function of the center.

32%	12 systems	100% dependent on other sources

32%	12 systems	Partially dependent on other sources
36%	14 systems	Perform all document input in-house

The 12 systems that receive 100% of their document input from other sources are computer systems. They use one of three approaches:

Receiving magnetic tape of document records	(9 systems)
Having dissemination centers perform SDI	(2 systems)
Using only author-indexed documents	(1 system)

The 12 systems partially dependent on others for their input are divided as follows:

Receive some input from dissemination centers (6 systems)

Use contractors for all indexing or whole SDI process (6 systems)

A little less than two-thirds of the systems surveyed rely on other sources for at least part of document input. Acquiring ready to use document input seems to be an advisable approach if the documents are compatible with user information requirements. Receiving document input from dissemination centers often simplifies the task of designing and implementing a functional SDI program because some procedures will be indirectly controlled by the center. The format of document input may have some effect on user profile construction and matching techniques. Software packages for computerized systems and consulting are often available from dissemination centers.

Question 2. Average document input time.

Input time refers to the time it takes to prepare document records so that they are ready to be matched against user profiles. As we indicated in the preceeding section, some systems do no document input processing. The following summary of responses clearly shows the time saved by using dissemination centers or contractors.

11 systems have no document input time

19 systems take between 2 and 90 minutes per document*

1 system requires 6.5 hours per document

6 systems made no response

*NASA's two systems are counted as 1, since both systems use same document input.

There are seven non-computerized systems in this survey. Four of them have measurable document input time and are scattered evenly through the distribution. One is the second fastest, and another is the second slowest. Hence, type of equipment seems to have little direct relationship to speed of document inputting.

We looked at two factors that we anticipated would be related to input time: document volume and depth of indexing. We found that no relationship between document volume and input time exists because there is no measurable increment between time and volume nor even a fairly recurring pattern of increase in time versus a decrease in volume. We had only slightly better luck with depth of indexing versus input time. These factors are related in 50% of the systems, but no increment could be found.

We concluded that document input techniques are highly individualized. The only significant group is the group of ten systems (30%) that do no document input processing.

Questions 3 and 4. Number of docu-

ments processed into SDI each month and the portion they are of total documents reviewed. Percent of documents inputted into SDI.

Almost all systems (28 out of a measurable 34) process 75% or more of the documents that are reviewed for SDI. Those systems receiving total document input from another source accept 100%, since no reviewing or purging is done by any of those particular systems surveyed. Crerar Library inputs only 10% of reviewed documents, probably because all new acquisitions come to the information specialists' attention. The Bureau of Reclamation accepts only 5% of the documents reviewed in order to input only the most significant documents into SDI. The indexers at Fort Belvoir select about 25% of the documents for SDI, and the rest are compiled for a booklet of recent acquisitions that includes document index terms.

Number of documents processed.

The range of the number of documents processed into SDI per month is very wide, from 25 documents to 30,000. The systems were ranked in order and divided into two groups. Group I contains twenty systems and has a range from 25 to 1200 documents per month. Group II has fourteen systems that range from 2200 to 30,000 documents. Four systems were unable to supply data for this question. See Figure 2.

We tested three factors that we assumed would affect the number of documents a system processes: the number of documents per user or profile, the number of SDI runs per month, and the number of documents per user or profile per run. These factors were analyzed for each of the thirty-four systems to determine whether the two arbitrarily determined groups were essentially the same: that is, perhaps each system processes approximately the same number of documents per user during each SDI run. The mean averages of tested factors appear below:

	Group I	Group II
Doc/mo	473	7592
Profiles	336	554
Doc/Profile	11	86.9
Doc/Profile/ Run	7.1	37.0

After reviewing the figures above, we excluded the two largest systems from Group II because they were much larger than the other systems, but only insignificant changes appeared in documents per profile and documents per profile for each run.

The systems in Group II not only have more documents and more profiles, but they process more documents per profile. We looked at the systems that form these groups to explain the difference between them. Group I contains most of the experimental and partially operational systems. These systems often work with a small number of documents in order to establish indexing procedures and matching techniques. Group II contains commercial SDI systems (Crerar and Aerospace Research Applications Center) and NASA. IBM's corporate-wide SDI system and Ames Laboratory are also in this group. Even though Ames has a relatively small user population, it has the largest document input, partly to establish the feasibility of a centralized dissemination center not limited in subject coverage.

We concluded that an SDI system can be established to serve nearly any size group with nearly any volume of document input. No ratio of documents to users exists for the surveyed systems. Non-computerized systems were evenly divided between Group I and Group II; hence type of equipment or lack of equipment has little effect on size of the system. Each system, whether large or small, seems to be able to fulfill its function,

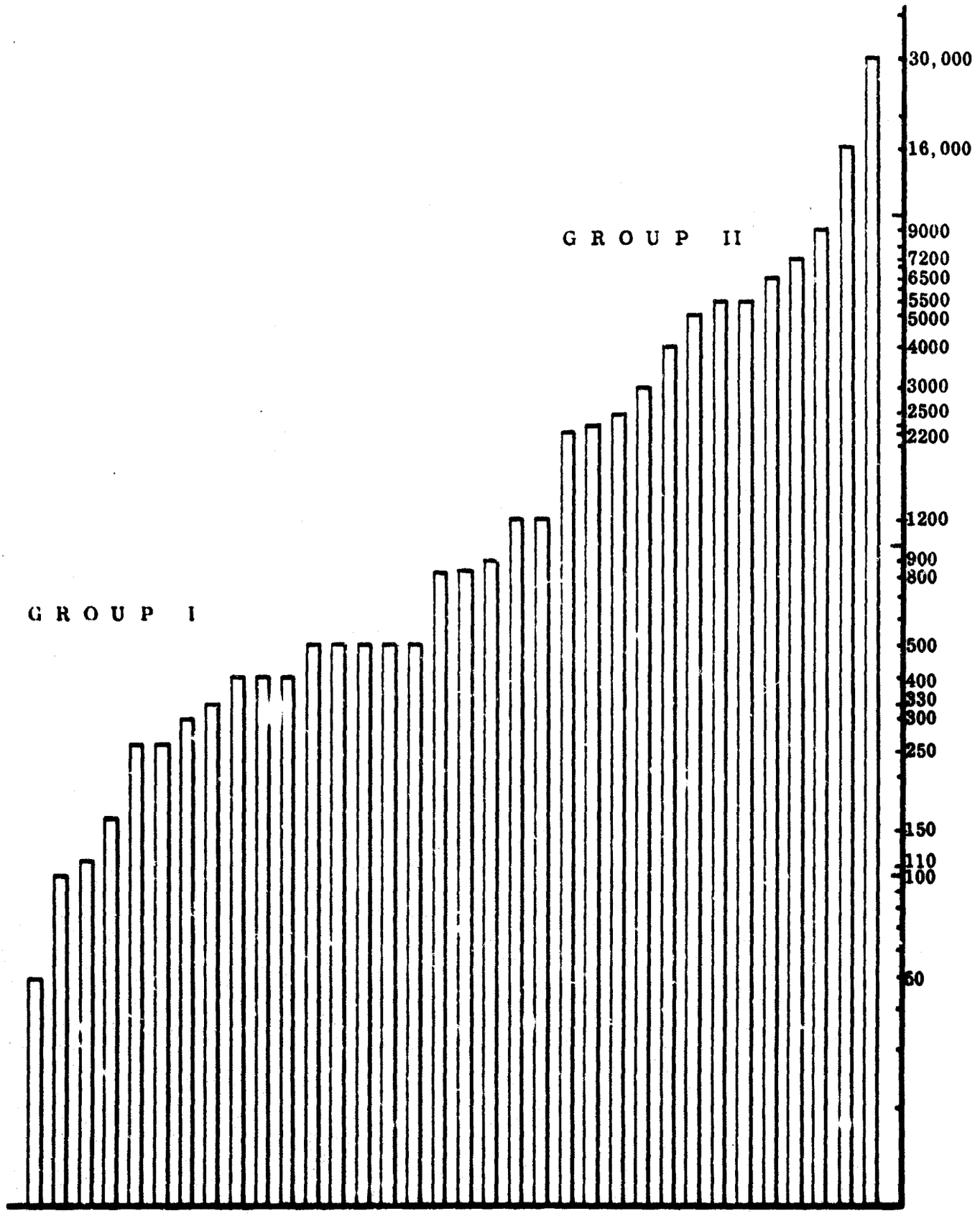


FIGURE 2. Number of Documents Processed per Month

regardless of the ratio of documents to users.

Questions 5,6 and 7.

Indexing procedures: vocabulary and maximum, average, and minimum number of index terms per document.

Participants were asked whether they use free language or a controlled vocabulary to index documents.

21 systems use controlled vocabularies

20 use internally developed vocabularies

1 uses external vocabularies

7 systems use free language

10 systems subscribe for total document input; hence the questions are not applicable

Use of Controlled Vocabulary

The ten systems that subscribe to dissemination centers for total document input have no control over document inputting methods. Hence, these systems are excluded from analysis of indexing procedures. However, we found that the centers supplying document input used controlled vocabularies in seven cases. The other three are participants in Chemical Abstracts Service SDI experiment, and for this experiment Chemical Abstracts uses free language abstracts as document input.

The majority of SDI systems surveyed use controlled vocabularies. Matching technique and profile construction are simplified when controlled vocabularies are used. Twenty of the systems developed vocabularies tailored to their needs either by modifying or supplementing pertinent external vocabularies or by constructing completely original vocabularies. The only system relying on external vocabularies, the U.S. Army Natick Laboratories, used four thesauri for a nine-month pilot test. The pilot test has been completed, and procedures are being analyzed in order to implement a fully operational system. During this time a vocabulary is being constructed using the four thesauri as a base. Hence, all surveyed systems that use controlled vocabularies establish internal vocabularies.

Seven systems use free language. Four of these use computers and three are manual. Ames Laboratory SDI system acquires all document input from other sources. Ames uses several sources but needs only one program to input document information regardless of vocabularies and formats used by Ames' sources. As a result document input includes all of the following tyras of format: title and author only, indexed document citations, abstracts and full text. One other computerized system uses free language abstracts as document input (Bonneville Power Administration). The corporate-wide IBM SDI system inputs free language text. One other computerized system indexes documents using free language.

Three systems using free language are manual systems. Documents require no indexing in order to be matched against user profiles. Matching documents and users requires judgement of information specialists who understand user needs but are not limited to a fixed vocabulary or to a formal search strategy.

Vocabulary Size

Most vocabularies are between 1500 and 9500 terms. The average number of index terms per document is usually 20 or less. We found no relationship between the size of vocabulary and the average number of index terms nor between volume of documents processed and the average number of index terms.

In summary controlled vocabularies are preferred to free language. Internally developed vocabularies are or will be used by all systems using controlled vocabularies. Non-computerized systems are divided between controlled vocabulary and free vocabulary with four in the former and three in the latter. Most controlled vocabularies contain between 1500 and 9500 terms. Although most of the systems that perform their own document inputting functions indicate that there is no upper limit to the number of index terms per document, only three systems and Chemical Abstracts use abstracts or text as input. Most systems index up to twenty words per document.

Questions 8 and 9. Type (internal or external/one-time publications or journal) of document input.

Participants were asked to indicate what portion of SDI input is one-time publications and what portion is selected from journals or periodicals. These two choices combine to describe total document input. Under each division participants indicated how many one-time documents and journals are internally or externally generated documents. One is a corporate-wide system; hence documents written by anyone in IBM or found in IBM journals are internal. Another system serves only the installation of Owego; hence only documents written in-house are internal. The third system is staffed by and serves two IBM offices at Poughkeepsie and Kingston. Hence documents written by the user community, either at Poughkeepsie or Kingston, are internal.

The ten systems receiving ready-to-run document input are excluded from supplying information on type of document input.

4 systems are fairly evenly divided

24 systems rely on either one-time publications or journals and periodicals for all or most of document input.

13 systems use one-time publications

4 use external documents

7 use external documents

2 are fairly evenly divided

11 systems use journals or periodicals

11 use external journals or periodicals

Two trends become apparent. SDI systems use externally generated documents. Most of the surveyed systems are also seen to rely on a single document type. The implication here is that systems find one or just a few sources of document input that best fit users' needs. We think it is fair to assume that significant information is likely to be presented in all four forms, although internally-generated journals are probably rare. However, only four systems present a mix of one-time publications and journals. By narrowing the number of sources, systems can process document input with less work.

Question 11. General subject areas of document input.

A chart has been constructed to indicate the subject areas of SDI for each system. (See Table 1.)

4. Profile

The SDI subscriber's interest profile, i.e., the aggregate of unit terms describing the user's areas of interest, is one of the very few features borne in common by all SDI systems being, as it were, one of the defining criteria of this type of information specific activity.

This section is devoted to an exam-

	Aerospace Topics	Chemistry & Pharmacology	Computer & Information Sciences & Communication	Earth and Ocean Sciences	Electronics	Engineering	Life Sciences	Materials, Methods & Equipment	Mathematics	Military Topics	Missiles	Nautical Topics	Nuclear Topics	Physics	Company Products	Client - Centered
1. U.S. Army Electronics Command	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
2. U.S. Naval Weapons Laboratory	X	X	X		X	X	X	X	X				X	X		
3. U.S. Navy Bureau of Ships				X						X		X				
4. Search Information Center, John Crerar Library																X
6. American Cyanamid Company, Pearl River		X														
7. Applied Physics Laboratory, Johns Hopkins University	X		X	X	X	X	X	X	X	X	X	X	X	X	X	
8. Bell Telephone Laboratories			X				Psych.		X							
9. U.S. Army Natick Laboratories		X		X	X	X		X		X						
10. Aerospace Research Applications Center			X	X	X	X	X	X	X				X	X		
11. Carter - Wallace, Inc.		X					X								X	
12. Nuclear Safety Information Center, U.S. Atomic Energy Commission	X				X		X	X					X	X		
13. VITRO Laboratories					X	X				X	X	X				
14. Fort Detrick Technical Information Division		X					X									
15. Harry Diamond Laboratories, U.S. Army Research Office						X			X	X						
16. Naval Reactor Information Center, General Electric Co.						X						X	X			
17. Bonneville Power Administration			X	X	X	X								X		
18. U.S. Bureau of Reclamation				X		X	X	X	X							
19. American Cyanamid Co., Bound Brook		X														
20. Ames Laboratory, U.S. Atomic Energy Commission			X		X	X			X				X	X		
21. Health Research, Inc., Roswell Park Computer Center		X					X									

	Aerospace Topics	Chemistry & Pharmacology	Computer & Information Sciences & Communication	Earth and Ocean Sciences	Electronics	Engineering	Life Sciences	Materials, Methods & Equipment	Mathematics	Military Topics	Missiles	Nautical Topics	Nuclear Topics	Physics	Company Products	Client - Centered
22. Eastman Kodak (Engineering Index SDI)					X			Plas-tics								
23. Eastman Kodak (Chem. Abstracts SDI)		X														
24. Western Electric Co.		X	X		X	X		X	X					X		
25. Eli Lilly & Co.		X					X									
26. National Cancer Institute		X					X									
29. Foreign Technology Division, U.S. Air Force	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
30. IBM, Owego, New York	X		X		X	X								X		
31. IBM, Armonk, New York			X		X	X			X					X	X	
32. Case Western Reserve University		X					X									
33. Martin Marietta Corporation	X				X					X	Air			X		
34. IBM, Poughkeepsie and Kingston, New York			X			X			X					X		
35. U.S. Army Mobility Equipment Research and Development Center		X		X	X	X		X	X	X				X		
36. Douglas Aircraft Company	X				X	X	X			X	Air					
37. B.F. Goodrich		X						X							X	
38. M. & T. Chemicals		X						X							X	
39. Aero Jet General Corporation	X															

ination of various profile techniques of generation and control. The manner in which the system internally operates on these profiles is examined in a later section (matching).

Vocabulary

In the selection of the unit terms to be contained in his interest profile, the user is allowed free choice in less than one third of the systems surveyed. The great majority of systems impose some manner of controlled or prewritten vocabulary from which the user must make his choice. It is of interest to note that in the free choice situations, over half of the systems actually operate on the user's selected terms. In the remaining systems, the user-selected terms are translated into some form of controlled vocabulary that the system uses internally.

Four computerized systems use free language document input and profiles. The ease and flexibility of document inputting in free language computerized system is somewhat overshadowed by the complexities of profile construction. The user may include any term he wishes and long profiles or multi-profiles are encouraged. However, the user also has an obligation to include alternate spellings and forms, e.g., ionization, ionisation, Charles R. Sage, C.R. Sage, symbolic and alpha chemical formulas. Synonyms and related words need to be included also and are generally the user's responsibility. However, SDI personnel are particularly aware of the profiling difficulties and are prepared to give additional aid.

The significant majority of systems operates, at least internally, on some form of controlled vocabulary, or thesaurus. These systems in turn are divided with two-thirds using internally generated special purpose thesauri and one-third using a published thesaurus such as D.D.C. (the Defense Documentation Center vocabulary), MESH (the medical vocabulary of the MEDLARS system), or E.J.C. (Engineers Joint Council thesaurus of engineering terms).

Construction of Profiles

Only one of the systems surveyed practiced "autoprofiling;" i.e., constructing the user's profile automatically from the user's job description, project description, library request records, etc. This concept, a popular theory in the past, has apparently not gained widespread acceptance in practice.

Constructing profiles regardless of complexity is a difficult, time consuming task. One of the participants indicated that most of the time users do not know what to ask for to produce desired results. Some systems have employed additional means for making profiles operational more quickly. One experimental system performs matching according to each user's profiles but sends all abstracts to the users for comments on all for the first few SDI runs. One system asks users to bring prototype documents that reflect user interests. In addition to profiling aids most systems expect staff members to spend time aiding users in profile construction.

Three surveyed systems have removed the profiling task from the users altogether by presenting the user with defined subject categories each broad enough to precipitate notices. After the user makes his selection, systems personnel construct profiles. One commerical SDI system offers users individually tailored profiles and this new approach to profiling. In some cases the user's cost is cut in half by picking subjects, and there has been a 70% change-over from individually tailored profiles. The system perhaps more graphically than any other example illustrates that users believe their information needs can still be met with this approach. One IBM system, called SCAN (System 34), is the forerunner of these systems, having been in operation

for three years.

Terms per Profile

As to the profiles themselves, a little more than half of the systems place no restrictions on the number of terms a user may have in his profile. In these systems with limited term profiles, the upper limit varies from 7 to 9999 terms. A qualification of those systems with very high upper limit figures (1000 or more terms per profile) appears when they are referred back to the type of vocabulary being used. Systems permitting open (or free) choice vocabularies either have very high upper term limits or no limits at all. Controlled vocabulary systems, on the other hand, tend to produce very definite limits with a range from 7 to 1000 terms with most profiles (for this type of system) containing 200 terms or less. For all systems where it is possible to compare the average number of terms per profile against the maximum, the average number of terms is less than half. For systems with no upper limit, the average is 30 terms per profile or less. Only two systems average over 100 terms per profile, and in both cases the average is less than half of the maximum. Since profiles average considerably fewer terms than the maximum, the systems are designed with larger capacities than are currently being used.

It is difficult here, as in the previous sectional analyses, to draw very many revealing conclusions. A tendency toward highly individualistic designs becomes increasingly apparent with the exception of the understandable system preference for established or "fixed" vocabularies.

Use of Weights and Logic

An examination of the systems' internal mechanics as regards profile manipulation, on the other hand, produced several marked uniformities among the study population. Over two thirds of the systems utilized some form of weighting or syntactical logic between terms within a user profile. These techniques varied from a simple set of logical operators (and, or, not) to scales of interger weights with ranges as high as 1000. Those systems not utilizing weighting or logic treat each term in the profile as having equal importance to the user and determine relevancy or degree of interest by the number of terms matched against a document. See the discussion of matching in Section 6 for a detailed discussion of matching techniques.

5. Profile Modification

Most or all of the SDI user population will, over a period of time, experience changes in its areas of technical interest. Thus, to maintain a relevant service, the SDI system must be flexible in its ability to alter a user's interest profile. This section of the study is directed at a review of the methods the various systems employ to adapt to changing user interests.

All of the surveyed systems permit the alteration of user profiles. This appears to be a uniform feature of all SDI services, and variations are encountered only in the means and frequency of modification.

Over 65% of the systems studied altered profiles or permitted profile alteration as the result of each distribution of user notices. Of these systems, those in the experimental or developmental phase relied heavily on personal interaction (interviews) between the user and system's staff. Those systems in full operation, on the other hand, relied almost entirely on some form of routine notice by the user, usually a returnable portion of the user notice.

Approximately 20% of the systems only allowed alteration of profiles at fixed calendar intervals. These periods ranged from bimonthly to semi-annually.

The remaining systems (approximately 12%) yielded a variety of responses to the question of frequency, although the implication is clear that modifications are accepted at the request of the user. These systems may be regarded as minor variations of the set which permits modifications per notice.

Interestingly, only one system (Ames Laboratory, USAEC) indicates that the profile is modified automatically by the systems computer program. This technique, as with autoprofiling, is another instance of a popular theoretical concept which has experienced only limited practical acceptance.

6. Matching

Matching is the focal point of SDI. Profile construction and document input are done in preparation for matching users' interests against new documents. As a result of matching, users receive notices of relevant documents. Two questions are discussed here:

Question 1. Method of matching against documents

Question 2. Condition necessary to send a notice to a user

Question 1. Method of matching profiles against documents.

The four combinations of matching user profiles against document profiles were offered: individual or aggregated user profiles against individual or aggregated document profiles. "Aggregated" means that all terms are reordered into a single master list so that no term appears more than once. Using aggregated profiles takes less space in computers and in general simplifies linear matching. However, creating a master list necessitates sorting operations to get terms associated with the proper user or document and to remove duplicate notices for the same document.

The participants use the following approaches:

24 Individual user profiles against individual document profiles.

1 Aggregated user profile against individual document profiles.

3 Individual user profiles against aggregated document profile.

4 Aggregated user profile against aggregated document profile.

6 Not applicable.

The question is not applicable to four manual systems that use SDI personnel to determine hits. (The other three manual systems have formal matching strategy.) The other two systems for which this question is not applicable do not perform SDI functions in-house.

One of three matching strategies are used by most systems: linear matching (14), boolean logic (6) and weighted terms (9). Linear matching can be performed efficiently using any of the four approaches listed above, but most linear systems prefer the first choice. The first choice seems almost a must for systems using boolean logic and weights, and most of these systems do use individual user and document profiles. However, two systems using weights aggregated document records, and one of these systems aggregated both.

Question 2. Conditions Necessary to Send Notices to Users.

In the matching process it is necessary to determine when a match between user and document is important enough to send the user a notice. This level of significant match is called the threshold or hit level. For instance, if the requirement is that at least three terms in a user's profile match terms in a document profile, the hit level is 3.

In order to look for meaningful re-

lationships, the systems are divided according to the type of user profile used: linear terms, boolean logic, weighted profile terms, and other. A tally of answers is presented in the chart below:

	Threshold Vary?		Controlled By		Varies By	
	No	Yes	User	System	User	Document
[illegible]	5	6	3	3	6	0
[illegible]	6	-	-	-	-	-
[illegible]	7	7	2	5	6	1
[illegible]	2	1	0	1	1	0

The same six systems excluded from the discussion of the preceding question are excluded here also. In order to see some of the different approaches given in the chart, representative systems will be explained.

Fort Monmouth uses linear matching. All user profile terms are independent and equal in value. Fort Monmouth allows users to use negative words that override matches. One match is sufficient for a user to receive a notice of the document.

Boolean logic uses the boolean operators and, or, and not to connect user profile terms into logic statements. The hit level is fixed. Although it varies from system to system, it does not vary within a system. Either the document terms satisfy the logic statement, or they do not. One interesting approach to boolean logic was developed at Fort Detrick. A boolean threshold of 10 is fixed; user profile terms are "weighted" in such a way that desirable combinations of terms sum to 10 exactly and a notice is generated. If the weights of the matched terms exceed or are less than, no match is made.

The SDI system at Ames Laboratory is an unusual system, since it is the only one that varies the hit level according to documents. In this system user profile terms are weighted between 0.0000 and 0.9999 to four significant digits. Most systems assign hit scores to user profiles, but Ames assigns a hit level to documents according to the depth of indexing. Texts have a hit level of 0.7500; abstracts are 0.5000; and titles are 0.3000. This system needs this sort of approach for efficient service because of the varied formats for document input.

Two systems in the "Other" group use a hierarchal indexing system from general to specific subject areas for both documents and user profiles. A match occurs when a document code matches or is more specific within the same area as the user term code.

The other system in this last group, Bonneville Power Administration uses weights and boolean operators together. This system is one of the three surveyed systems using abstracts or text as input. Using abstracts requires complex profile construction and matching technique. Bonneville uses a weighting scale for profile terms from -9 to -9 and a hit level from 00 to 99, but in addition other features are included. "Synonym families" simulate "or" in boolean logic. The user uses a code for those situations where one word is as good as another; but if both were to be found in the same document, they would count only once. "Not" in this system is not weighted and serves as total negation of a document regardless of other matches. "Must" is also not weighted and means that a document will be suppressed unless the must term is in the abstract. Weighted, not and must terms are used both for exact and root terms. Root terms match with all document words beginning with the same letters in the same spacing and punctuation.

The most significant factor in hit level variation is that if the hit level varies, it varies user by user rather than document by document. When linear matching is used, the systems

surveyed are fairly evenly divided between the several options. Boolean strategy implies a fixed hit level. When weighted profile terms are used in matching, the hit level usually varies and it varies user by user. One system using weights adjusts users' profiles so that all have a hit level of 0.999. Each user constructs his profile using a weighting scale from 1 to 9, and he assigns the proper hit level. Then the system sets the hit level at 0.999 and adjusts the weights within this structure.

7. Notice

According to the SDI concept, SDI is an alerting service. The point of contact between the system and the user is the notice of new documents of probable interest. Notices mark the end of the SDI process.

Questions 1 and 2. Content and form of user notice.

Question 3. Frequency of distribution.

Questions 1 and 2. Content and Form of user Notices.

Within system limits the notice is established to be compatible with users, needs. Since SDI systems have two factors in common - notice as only the point of interface between user and system, and notice design as the measure of user information requirements - an analysis of user notices should yield a typical user notice. Experimental and partially operational systems are excluded because the user notice is one of the parameters usually under development and hence does not truly reflect coordination of user needs and system capacity.

The typical notice contains an abstract and accession number. Systems including classified documents in the input print citations for classified documents and abstracts for unclassified material. If the SDI system uses weights, a weight factor may also be included. A return request form is included, implying that each installation possesses the document is cites. Some method for profile modification may also be included.

Although the content of a notice can be generalized, the form used by the surveyed systems is varied. The most common characteristic is that the notices are machine printed. Since most of the systems in the survey are computerized, printed notices are expected. However, half use punch cards or double hinged punch cards. Some print on computer paper, and some use cards of other than punch card size. One system uses machine print-out on cards for all information except the abstract, which is xeroxed onto the cards.

Only three fully operational systems deliver documents as a notice. An earlier IBM experimental SDI system distributed copies of the document; but as the number of users increased, distributing documents became economically inadvisable. In general user comment did not endorse document distribution. A user seems to prefer an abbreviated version of the document, particularly when a system has a large data base. Sometimes just an awareness of the document is sufficient so that an abstract is satisfactory. However, the function of each system should be considered before final judgment on document delivery as the notice is made. The three systems that send documents initially are designed in such a way that not to deliver documents is impractical.

Question 3. Frequency of Notice Distribution.

The twenty-five operational systems distribute notices in the following manner:

5 Daily
9 Weekly
6 Twice Monthly
5 Monthly

Frequency of notice does not seem to be related to documents per run.

Budget figures are available for only ten systems in this group. For these systems there is no relationship between money spent on SDI and the number of runs made. Frequency of notice distribution depends on user's need for current information and systems design rather than on budget or document volume.

8. Equipment and Personnel

Equipment

Utilization of fixed resources, as with the budgetary data, was difficult to define in the survey responses. The principal reason is that all but a few of the SDI systems studied were peripheral functions of larger systems, sharing with the parent system its equipment, facilities and personnel.

Proportions of equipment and personnel devoted to SDI functions not only varied widely from system to system, but within systems from time to time as requirements and work loads varied.

All of these factors combine to defy any revealing comparative analysis and to obscure correlations between various other system factors such as volume, population and cost.

Of a number of quantifiable factors, we found that approximately 75% of the systems surveyed utilized a computer capability in one of numerous forms and levels of utilization.

In those systems reporting personnel assignments, the large majority (again over 75%) used six or less persons as permanent staff. Again it was not possible to make any correlation between man-hours and budget. In fact, attempts in this area lead us to believe that at least one of these two factors has been reported erroneously in numerous instances.

Physical configurations varied throughout a wide range of sophistication from manual card files to an IBM system 360/50 with a fully loaded channel. Methods of source data automation centered strongly around the use of keypunching with two instances of the newer technique of keytape entry.

One of the interesting aspects of equipment utilization is the sharp demarkation found when progressing from manual to automatic systems. There is virtually no middle area of semi-manual or semi-automatic process (i.e., punched cards, optical coincidence, etc.) while the remaining were either entirely manual or fully automatic.

Another interesting aspect was the relatively high occurence of multiple computer utilization. Fully 33% of the represented systems used two or more computers in various processing steps of the SDI cycle. Two of the systems indicated the use of three computers, and one system listed four. In the great majority of dual computer use, the second machine was of a relatively restricted capability (such as IBM 1401, Univac 1004) and used primarily as an off-line printer. Diffusion of processing steps among two or more computers in the same system of roughly the same capability level was primarily to permit utilization of existant software, developed for or shared by other non-SDI applications.

The above summary, and particularly the last observation, indicates strongly that SDI is a relatively low priority operation in a multipurpose computer installation. In fact, at this point in the survey, the SDI concept is emerging as a precariously established system, developed on an ad hoc basis and existing at the sufferance of higher priority budgetary and resource requirements.

Personnel

Personnel by category or professional level were reported in very nearly

all systems studied and permitted still another interesting observation. In its earlier form, SDI was viewed as largely a semi-professional or clerical occupation. However, a profiling of personnel engaged in this activity shows that for each clerk there is one technician (indexer, abstractor, etc.) and 1.25 full professional analysts. On the other hand, there are only 0.5 persons engaged in machine operation and 0.2 persons listed as programmers.

Analyzing these profiles by system type (e.g. computerized or manual) shows a slight shift in the clerical based ratio, as indicated in the below table, which appears to indicate that the introduction of a computer not only introduces new personnel types (programmers) but seems to increase the overall staff as well.

	All	Computer	Manual
Professional Analyst	1.25	1.4	0.75
Technician	1	1.1	0.4
Programmer	0.2	0.2	0
Machine operator	0.5	0.6	0.25
Clerk	1	1	1

The temptation to believe that this phenomenon explodes the myth of the computer as a labor saving device must give way to plausable explanation. The increase in the ratio of professional analysts, for example, simply reflects management's reluctance to leave a sophisticated and higher cost system in the hands of technical or semi-professional personnel. The rise in machine operators reflects the addition of such occupations as keypunch/keyverifiers, and the rise in technical staff follows the computer's ability to process large volumes. A more meaningful analysis of course would be the ratio of staff to users but again no correlations are obtainable from the surveyed data.

B. THREE SDI SYSTEMS IN THE UNITED KINGDOM

One feature of this state-of-the-art survey is to include foreign SDI systems. Unfortunately documentation on SDI systems outside the United States is not as numerous as that for U.S. systems. The time necessary to do a thorough search was unavailable; hence the three systems included in this survey should not be considered as a representative sample. We are aware of one more SDI system serving a steel industry based in Glascow, Scotland, and cited in ASLIB Proceedings, January 1965. This manual system is large. Unfortunately, we were unable to send a letter to this company to invite participation in the survey.

The three systems that were contacted responded immediately with a completed questionnaire and very helpful information. Their thoughtfulness and promptness is greately appreciated. These three systems represent a cross-section of types: one commercial system, one government system, and one research and development system.

The Scientific Documentation Centre based in Dunfermline, Fife, United Kingdom is the commercial SDI system. The system has a particularly large document volume, 300,000 items yearly, from a wide range of sources and countries covering a wide range of topics. Documents include all of U.S. Government Research and Development Reports and appropriate parts of U.S. and British Thesis Titles. The system started four years ago solely for SDI. Gradually the topics are being made available for retrospective searches also so that half are included in a storage bank at present. Science, technology, medicine, botany, zoology, and engineering are some of the fields included. Fees vary according to topics selected. Most topics cost between fifteen and fifty pounds per year. Many of these prices are less than American systems.

A potential user reviews the list of topics available and indicates his choices. The staff at the Centre construct the profile. The user receives weekly notices on

3 x 5 card or a punch card according to his preference. Each card contains a citation and keywords. The user may order the document, but ordering documents requires an additional fee.

Specific questions on users and profile construction are considered confidential, as are the workings of the system. A staff of forty, mostly students who work part-time, maintains this manual SDI service, once again establishing that manual systems are not limited in size or capability.

The Culham Laboratory, United Kingdom Atomic Energy Authority, London, operates an SDI system both for internal and external users. The lab has always supplied a manual SDI service and until three years ago, it was satisfactory. However, the number of users increased until it became impossible for the information staff to meet the needs of all users adequately. Experiments were conducted that resulted in a fully operational system being implemented a year and a half ago. This system operates with no additional staff, a minimal amount of additional clerical support, and no interruption or diminishing of regular library functions.

This system uses document titles as input. When titles are not sufficiently descriptive, additional words are used. The average number of terms per document is twelve. Profiles are constructed in the form of a matrix with ten columns and thirty rows, allowing a maximum of 300 words. The average profile, however, contains only 20 terms. The terms are connected by and, or and not. This SDI system operates without using a controlled vocabulary.

The computer used is an English Electric-Leo-Marconi KDF9 with a random access disc file of about four million words. Core storage holds 32,000 words and magnetic tape files contain about 2 million words per reel. The input device is an ULTRONIC 811, a punched paper tape machine.

Personnel involved in this system's operation are particularly pleased with the number of relevant notices each user receives, particularly because only titles and authors are used as input. One factor involved in this high precision ratio is that prior to sending notices to users the information staff reviews each user's notices to eliminate irrelevant citations.

There are 200 users, 10 managers and 190 research scientists, who receive notices twice weekly if users are in-house and once weekly if the users are external. Document flow is 150 per week or about 650 per month. Input time per document is 5 minutes and includes preparing the paper tape. The subjects included in SDI are plasma physics, controlled thermonuclear fusion, ultra high vacuum, cryogenics, super-conducticity, lasers, microwaves.

The Laboratory is successful in effecting smooth changeover from manual to computerized SDI service without overburdening the staff or jeopardizing other library functions.

The third SDI system was developed as a project to test the value, cost, and acceptance of a selective dissemination of information system. The project is expected to last four years. For the first two years the National Electronics Research Council was in charge. The primary task was to establish a vocabulary and indexing procedures. The vocabulary, patterned after the Engineers Joint Council thesaurus of engineering terms, contains 1800 words and grows at the rate of approximately 80 terms per week. That rate of growth will sharply diminish after the system has been operational for awhile. The project has been transferred to the Institute of Electrical Engineers who received a grant of 61,350 pounds for 27 months to continue the project. IEE's task is to select 600 electronics research workers from universities, industries, and government and construct users' profiles. In the last year of the project the system is expected to be fully operational and to distribute

Group I

	A	B	C
Budget per year	$4500	$48,000	$20,000
Users [Profiles]	15 [?]	38 [38]	5 [11]
Documents per year	6000	48,000	4800
Runs per year	26	6	12
Cost per user [Profile]	$300 [?]	$1265 [$1265]	$400 [1818]
Cost per document	$0.75	$1	$4.17
Cost per run	$175	$8000	$1667
Documents per run [per user]	230 [15]	8000 [205]	400 [80]
Cost per user per run	$12	$210	$333

Table II. Experimental and Partially Operational Systems

Group II

	A	B
Budget per year	$3000	$8000
Users [Profiles]	114 [114]	103 [103]
Documents per year	10,000	1500
Runs per year	26	26
Cost per user [Profile]	$26 [$26]	$78 [$78]
Cost per document	$0.30	$5.33
Cost per run	$115	$308
Documents per run [per user]	385 [3.4]	58 [0.6]
Cost per user per run	$1.00	$3.00

Table III. Small SDI Systems

	A	B	C	D
Budget per year	$10,000	$25,000	$30,000	$30,000
Users [Profiles]	400 [400]	260 [260]	800 [1500]	68 groups [386]
Documents per year	60,000	3000	4000	6000
Runs per year	24	12	26	52
Cost per user [Profile]	$25 [$25]	$96 [$96]	$38 [$20]	$440 [$78]
Cost per document	$0.15	$8.35	$7.50	$5.00
Cost per run	$415	$2085	$1155	$575
Documents per run [per user]	2500 [625]	250 [.9]	155 [0.2]	125 [2]
Cost per user per run	$1	$8.00	$1.45	$8.45

Medium SDI Systems

Group IV

	A	B	C	D	E
Budget per year	$45,000	$50,000	$65,000	$130,000	$180,000
Users [Profiles]	220 [220]	12 industries [12]	200 [700]	2000 [2000]	700 [?]
Documents per year	348,400	86,400	30,000	625	22,000
Runs per year	52	52	12	12	220
Cost per user [Profile]	$204 [$204]	$4165	$325 [$93]	$65 [$65	$255 [?]
Cost per document	$0.13	$0.58	$2.15	$218.22	$8.18
Cost per run	$865	$960	$5415	$10,835	$820
Documents per run [per user]	6700 [30]	1660 [140]	2500 [12]	52 [0.03]	100 [0.1]
Cost per user per run	$3.90	$80	$1.60	$5.40	$1.20

Table V. Large SDI Systems

weekly notices. The computer selected for this system is an English Electric KDI 9.

Early in 1968 the first SDI runs will be made. At the time of contact many questions were unanswerable because procedures will be determined by practice. The indexing and profiling vocabulary and indexing techniques are fairly well established. However, the vocabulary is expected to grow rapidly for awhile. Document volume is expected to be 240 documents per week or about 1000 per month. However, input time, percent of documents accepted for SDI, and the average number of terms per document are unknown. Each user provides a written statement of interests, and staff members translate it into a user profile with an upper limit of 100 terms connected by logic operators. The average number of terms is unknown. Profiles will be modified according to each user's assessment of his notices. He will receive a citation and index terms for each relevant document. The staff and the computer are available, but the amount of time that will be spent on SDI function is unknown.

Since this project is a test of selective dissemination, much more time can be spent analyzing all the parts of an SDI system than most installations can allow. The final operating cost and techniques of this system will certainly be interesting and worthy of in-depth study after the system becomes fully operational.

C. COST ANALYSIS

We found that interpretation of SDI cost varies from system to system, primarily because SDI is a subsystem. For instance salaries of staff, mailing or distributing notices, cost of duplicating documents or appropriate parts, and cost of document acquisition are some of the factors that may be included as part of the cost; but whether they are included depends on each installation's budget policy and interpretation of SDI.

Fourteen systems in the survey supplied a yearly budget figure. The systems were divided into four groups that yeild some measure of compatibility within each group. Most of the experimental and partially operational systems are grouped together because document volume and users are considerably fewer than each system plans to include in a fully operational system. Hence relationships between cost and other factors are not stable. Small, medium, and large or commercial systems are the three remaining groups.

Group I. Experimental and Partially Operational Systems

System A	$ 4,500.
System B	$ 48,000.
System C	$ 20,000.

Group II. Small Systems

System A	$ 3,000.
System B	$ 8,000

Group III. Medium Systems

System A	$ 10,000.
System B	$ 25,000.
System C	$ 30,000.
System D	$ 30,000.

Group IV. Large and/or Commercial Systems

System A	$ 45,000.
System B	$ 50,000.
System C	$ 65,000.
System D	$ 130,000.
System E	$ 180,000.

The factors used in this cost analysis are document volume, number of users, and number of SDI runs per year. Tables listing the calculations used for cost analysis are found within the next several pages. Question marks indicate lack of information.

The three systems in Group I (Table II) illustrate that SDI systems begin with limited inputs. Two experimental systems have large enough SDI budgets to be in Group IV, and both probably will, once they become fully operational. One of these uses 38 of 500 eligible users. Currently the system uses National Library of Medicine MEDLARS tape. The magnetic tape, issued monthly, contains approximately 8000 citations. However, this installation makes an SDI run every other month using the tape for one month only. The system will eventually operate twice a month, and the sources for input will be increased to include other tape services. Hence the number of users will increase more than ten times; the number of SDI runs will increase four times; and the document level will more than double when the system becomes fully operational. The cost per user for each run, presently $210, will be reduced to approximately $5 if the budget remains constant.

Another system in this group is running an experiment to establish an alerting service for those involved in diabetes research. The source is fixed; a selection from MEDLARS' is being used. The document volume, 400 documents per run, reflects the documents selected for the five use[illegible]ather than the documents found in the Diabetes Literature Index, the portion of MEDLARS' citations. This system will become operational in a year. The cost per user is expected to be greatly reduced, since there is a large population available to use this service.

The cost per user for each SDI run at the third system in this group ($12) is much less than both of the other two experimental systems. However, the cost is high when compared to operational systems. This system is using only 15 to 400 eligible users and has selected 120 journals for the test collection.

These experimental systems do not spend money directly on the user as the fully operational systems do. Instead money is spent for testing approaches to document selection and indexing, profile construction and vocabulary control, matching technique, and for testing and debugging equipment. When a system is fully operational the cost per user is significantly less because the money spent for an SDI service is more directly channeled into sending notices to users.

Group II contains two small systems. One of these is a partially operational system but is included here because it is a stable perfected system. It is presently mannually operated; but after the system becomes computerized, the number of users will be increased. The computerized approach will operate in the same manner as the system does now. This system has done some careful cost analysis based on notices. The total cost for each summary is 20¢ and includes indexing, clerical operations and supplies. With the present number of users, indexing accounts for nearly half the cost. Indexing time should remain nearly constant; hence as the number of scientists increase, the cost per user will decrease. For the year's trial run about 6,000 summaries were sent. Each user averaged about 60 summaries a year, for a yearly cost per user of $12.00. Dr. Schneider,

system designer, expects to be able to deliver 1000 notices per year to each scientist for $100 (10¢ per notice) when the system has the proper number of users (Approximately 1000). The yearly cost per user we found in this survey, $78, is considerably higher than Dr. Schneider's calculated cost. We do not know where the differences arise.

The other system operates with the smallest SDI budget in the survey. Although this system has fewer users than most of the other ten systems in Group II - IV, it has more documents than half of them and more documents per user than six of them. This computerized system uses free language and indexes documents to a depth of 10 terms. No additional information is available to explain how this system operates. However, the cost per user for each SDI run is reasonable and compatible with other systems.

Group III contains medium size SDI systems. The document flow of System A (60,000 a year) and the cost per user for each SDI run of System D are interesting. Only two other fully operational systems equal or exceed the large document input per run of System A. This system subscribes to Defense Documentation Center for document input in the form of magnetic tape. The Center handles all defense and defense-related documents; hence document supply is large. The advantage of receiving ready-to-run document input shows in cost of the system. System D uses group profiles. The system serves the majority of the 600 engineers employed there by establishing profiles for 68 groups. The number of real users brings the cost down to approximately $1 per user, depending on how many people use SDI. A cost of around $1 per user for each SDI run is compatible.

APPENDIX A

SDI

QUESTIONNAIRE

(Revised for Computerized Systems)

Office of Scientific and Technical Information
Office of Aerospace Research
United States Air Force

I. GENERAL CONSIDERATIONS

1. What is the name of your organization? _____

2. Briefly how does your organization define the SDI concept? _____

_____

_____

3. What is the formal title of your SDI system? _____

_____

4. How long has your SDI system been in operation? ____ years ____ months.

5. What is your current budget for operation of your SDI system? _____

6. Is your system part of library or larger information center? _____

7. Is your SDI system

   ☐ fully operational

   ☐ partially operational

   ☐ experimental

8. If a copy is available, please forward a system manual. Please indicate if you would like it returned.

II. USER

1. Is population limit of the system

   ☐ Fixed?

   If yes, what is limit? _____

   ☐ Open-ended

   What is population size at present? _____

USER - continued

2. What is population composition?

   _____ managerial

   _____ other

3. How many people are eligible users? _____

4. What is the level of distribution? _____

   ☐ Individual

   ☐ Groups related by

      ☐ Discipline

      ☐ Administration

      ☐ Task

III. DOCUMENTS

1. Are any documents indexed before you receive them? _____

   If Yes

   a. From what source(s)? _____

      _____

      _____

   b. What percent of total documents are previously indexed? _____%

2. What is average input time (indexing, machine coding, and bibliographic citation) per document? _____

3. At what rate are documents processed? _____ per _____

4. Of the documents you review, approximately what percent are accepted by the system? _____%

DOCUMENTS - continued

5. Do you use

   ☐ Free indexing (without benefit of controlled vocabulary)

   ☐ Controlled index vocabulary

   a. What is its source?

      ☐ Internal

      ☐ External. Name _____

   b. How many terms are in vocabulary at present? _____

   c. At what rate do you add terms to vocabulary? _____

6. Is the number of index terms per document limited? _____

   If Yes, lower limit _____

   upper limit _____

7. What is the average number of index terms per document? _____

8. How many documents are one-time publications (technical reports, books, etc.)? _____

   How many one-time publications are internal*? _____

   How many are external? _____

9. How many documents are journals and periodicals? _____

   How many journals and periodicals are internal*? _____

   How many are external? _____

10. Is any indexing performed on a contract basis? _____

11. Please list general subject areas of documents included in your SDI system.

* Generated within parent organization

IV. PROFILE

1. Does user select his own profile? _____

   If Yes, does he use

   ☐ Free selection

   ☐ Controlled vocabulary

   What is its source?

   ☐ Internal

   ☐ External Name _____

   If No, briefly describe profiling process

2. How many profiles per user are accepted? _____

   How many profiles are now in the system? _____

3. Is the number of terms per profile fixed? _____

   If Yes, lower limit _____

   upper limit _____

4. Do you use a weighting system? _____

   If Yes, what is your scale? _____

   _____

5. Can the user override his profile and select automatic distribution of specific authors, journal titles, corporate authors, etc.? _____

6. Can user request scanning (monitoring) of specific sources? _____

7. What is the average number of terms per profile ?_____

V. PROFILE MODIFICATION

1. What is your system for profile modification?

   ☐ Returnable notice or portion of notice

   ☐ Interviews or questionnaires

   How frequently? _____

   ☐ Other. Please explain. _____

   _____

2. How often do you modify profiles? _____

3. Are profiles modified by

   ☐ Adding or deleting terms

   ☐ Modifying weights

   ☐ Both of above

4. If you use a weighting system, how are weights modified? (Include boolean log

   ☐ by the user

   If user changes weights, does he see

   ☐ list of matched terms

   ☐ list of document terms

   ☐ other _____

   ☐ by the system/please explain. _____

   _____

VI. MATCHING

1. Matching is performed by comparing

   ☐ individual interst profiles against individual document profiles.

   ☐ aggregated* interest profile against individual document profiles.

   ☐ individual interest profiles against aggregated* document profile.

   ☐ aggregated interest profile against aggregated document profile.

2. Does the threshold level (point below which no notification is made) vary? _____

   a. Who initiates threshold modification

      ☐ user

      ☐ system

   b. Is the threshold varied

      ☐ document by document

      ☐ user by user

   c. How is the threshold computed? Please explain briefly. _____

_____

* Combining all terms into a single master list of terms.

VII. NOTICE

1. Does user receive

☐ Citation	plus	☐ Weight or match factor
☐ Abstract		☐ List of matched terms
☐ Accession number		☐ Document index terms
		☐ Return request form
		☐ Profile modification form
		☐ Other

Please check all that are applicable.

NOTICE - continued

2. In what form is the notice?

- ☐ Machine print-out
- ☐ Pre-printed form (by machine or manual sources)
- ☐ Punch card
- ☐ Library card
- ☐ Document
- ☐ Microfiche
- ☐ Copy of table of contents
- ☐ Copy of cover
- ☐ Other _____

Please check all that are applicable.

3. How often do you issues notices? _____

Will you please attach a copy of your notice form if it is available.

VIII. INTERNAL SYSTEM CONFIGURATION

1. Components

List all central and peripheral EDP/ADP equipment used in the operation of your SDI system.

Type/Model	Time Allocated to SDI (in hou Input*	Matching**	Output***
_____	____	_____	_____
_____	____	_____	_____
_____	____	_____	_____
_____	____	_____	_____

2. Personnel

List all professional, technical, clerical personnel used in the operation of your SDI system. Indicate contract personnel where applicable.

Job Title	Time Allocated to SDI (in hours) Input*	Matching**	Output***
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____

* **Profile modification and document input**

** **Document - user comparison**

*** **Printing and/or issuing notices**

APPENDIX B

Participating Organizations

(Numbers were assigned for the survey
and appear throughout the report)

Government, Private Institutions, and University SDI Systems

Government: 14 installations, 15 systems

1. U.S. Army Electronics Command, Fort Monmouth, New Jersey
2. U.S. Naval Weapons Laboratory, Dahlgren, Virginia
3. U.S. Navy Bureau of Ships Technical Library, Washington, D.C.
9. U.S. Army Natick Laboratories, Natick, Massachusetts
12. Nuclear Safety Information Center, U.S. Atomic Energy Commission, Oak Ridge, Tennessee
14. Fort Detrick Technical Information Division, Frederick, Maryland
15. Harry Diamond Laboratories, U.S. Army Research Office, Washington, D.C.
17. Bonneville Power Administration, Portland, Oregon
18. U.S. Bureau of Reclamation, Denver, Colorado
20. Ames Laboratory, U.S. Atomic Energy Commission, Ames, Iowa
26., 27. National Aeronautics and Space Administration, Washington, D.C.
28. National Cancer Institute, National Institutes of Health, Bethesda, Maryland
29. Foreign Technology Division, Wright-Patterson Air Force Base, Ohio
35. U.S. Army Mobility Equipment Research and Development Center, Fort Belvoir, Virginia

Private Institutions: 21 installations, 22 systems

4. Search Information Service, John Crerar Library, Chicago, Illinois
6. American Cyanamid Company, Pearl River, New York
7. Applied Physics Laboratory Document Library found at Johns Hopkins University, Silver Spring, Maryland
8. Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
10. Aerospace Research Applications Center, Bloomington, Indiana

11. Carter-Wallace, Inc., Cranberry, New Jersey
13. VITRO Laboratories, Silver Spring, Maryland
16. Naval Reactor Information Center, General Electric Co., Schenectady, New York
19. American Cyanamid Company, Bound Brook, New Jersey
21. Health Research, Inc., Roswell Park Computer Center, Buffalo, New York
22.,23. Eastman Kodak Co., Rochester, New York
24. Western Electric Co., Princeton, New Jersey
25. Eli Lilly & Company, Indianapolis, Indiana
30. IBM, Owego, New York
31. IBM, Armonk, New York [corporate-wide system]
33. Martin Marietta Corporation, Orlando, Florida
34. IBM, Poughkeepsie and Kingston, New York
36. Douglas Aircraft Company, Santa Monica, California
37. B. F. Goodrich, Avon Lake, Ohio
38. M & T Chamicals Rahway, New Jersey
39. Aero-Jet General Corporation, Azusa, California

University: 1 installation, 1 system

32. Western Reserve University, Cleveland, Ohio

SDI Systems in the United Kingdom:**

33. Scientific Documentation Centre
34. Culham Laboratory
35. Institute of Electrical Engineers

*These systems supply SDI services to other industries, government agencies and universities either as part or all of their function.

** These systems have not been included in the statistical tabulation. They are discussed in Chapter I-F.

APPENDIX C

Detailed Tabulation of Data
[Responses to SDI Questionnaire]

The following several pages contain participants' responses to nearly all questions from the survey questionnaire. Some questions are omitted because they do not supply meaningful information. Explanations supplied by participants will be abbreviated and included.

Abbreviations used on the following pages are derived from the questionnaire, and it is hoped that they have been selected with care in order to be obvious when the questionnaire is referred to. Although the reader may experience some inconvenience while interpreting abbreviations, in the interest of conservation of space, it seemed as a best approach.

Some respondence failed to supply answers for three reasons: the information was private; the information was unknown; or a question was misunderstood and its answer was incorrect. Missing answers are coded with "?" regardless of the reason for omission. If the question is not applicable, then "N/A." will be recorded.

4. Time [in years]	5. Yearly Budget [in thousands]	7. Level Of Operation:
0.75	$10	Fully
0.67	Free + 2 clerical hrs/mo.	Partial
2.5	?	Experiment
15.00	$50	Fully
0.40	?	Experiment
4.00	$3	Fully
1.50	?	Partial
0.75	?	Experiment
5.00	?	Fully
4.00	$180	Fully
1.75	$30	Fully
1.75	?	Fully
2.40	$48	Partial
0.25	$4.5	Experiment
6.40	$25	Fully
1.25	$30	Fully
4.20	$130	Fully
0.40	$60 + computer rental	Experiment
2.80	$45	Fully
2.00	?	Fully
1.40	?	Experiment
0.40	?	Experiment
4.00	?	Fully
0.75	$72 + computer rental	Fully
3.75	?	Fully
1.00	?	Experiment
0.67	$8	Experiment
5.00	?	Fully
5.00	?	Fully
2.75	?	Fully
1.00	$60	Experiment
6.00	$65	Fully
3.00	?	Fully
2.67	?	F [SDI only]
5.50	?	Fully
0	0	F, inactive
17.00	?	Fully
4.50	?	Fully

Section II. User

	1. Fixed, open-ended population limit	2. Population Composition			1. Number of use
		Managerial		Other	
1.	Open	75		325	400
2.	Open		Both		300
3.	Open	10		0	10
4.	Open	12		0	12
6.	Open		Both		20
7.	Open	0		114	114
8.	Open		Both		800
9.	Fixed, 25 Users	0		25	25
10.	Open		Both		78
11.	Open	35		665	700
12.	Open	240		560	800
13.	Open	15		1385	1400
14.	Open	1		37	38
15.	Fixed, 40 Users		Both		15
16.	Open	65		195	260
17.	Fixed, 1000 Users	0		68	68 Contacts
18.	Open	100		1600	2000
19.	Open	4		72	76
20.	Fixed, 10,000 Users	30		190	220
21.	Open		Both		25
22.	Open	9		11	20
23.	Open	3		9	12
24.	Open	200		1800	2000
25.	Open	7		73	80
26.	Open	170		680	850
27.	Open	100		400	500
28.	Fixed, 103 Users	0		103	103
29.	Open	2%		98%	?
30.	Open		Both		195
31.	Open		Both		2800
32.	Fixed, 5 Users	0		5	5
33.	Open	10		190	200
34.	Open	65		65	130
35.	Open		Both		700 Users (225 Tasks
36.	Open	32		293	325
37.	Open		Both		Non-Operation
38.	Open		Both		?
39.	Open	5		95	100

	3. Number of eligible users	4. Profiles of Individuals (I), Groups related by discipline (GD), task (GT), Administration (GA)	
		Majority	Other
1.	2000	I	GD, GT, GA
2.	1000	I	
3.	50 Projects	GT	
4.	Open	G Industry	
6.	300	I	
7.	700	I	GD
8.	5000	I	
9.	800	I	
10.	Open	I, GD, GA	
11.	MS & Up	GD	
12.	4000	I	
13.	2900	I, GD, GA	
14.	500	I, GD, GT, GA	
15.	400 (40 for pilot)	I	
16.	1100	I	GD
17.	600	GT	
18.	2500	I	
19.	900	I	GD
20.	500 (determined	I	GD, GA G Industry
21.	by budget limit)		
21.	Open	I	
22.	600	I	
23.	600	I	
24.	2400	GD & GA (250 Profiles)	I (50 Profiles)
25.	775	I	
26.	Open	I, GD	
27.	Open	GD & GA	
28.	Open	I	GT
29.	?	I, GD	
30.	1500	I	GD
31.	?	I	
32.	Open	I	
33.	3000	I	
34.	130	I, GD, GA, GT	
35.	700 (225 Tasks)	GT	
36.	?	I	GD, GT
37.	600	I	
38.	?	I	
39.	3000	I	

1. External Source Input		% of document Input External	Internal	2. Input processing time per document
1.	DDC	100%	0%	N/A
3.	NASA/SCAN	100%	0%	N/A
3.	None	0%	100%	25 Min.
4.	None	0%	100%	2 Min.
6.	Chem Abstracts	100%	0%	N/A
7.	NASA	15%	85%	45 Min.
8.	Internal Author	100%	0%	5 Min.
9.	DDC	10%	90%	38 Min.
10.	Contractor	100%	0%	N/A
11.	None	0%	100%	67 Min.
12.	None	0%	100%	?
13.	None	0%	100%	N/A - Staff Decision
14.	MEDLARS	100%	0%	N/A
15.	Contractor	100%	0%	20 Min.
16.	None	0%	100%	25 Min.
17.	None	0%	100%	10 Min.
18.	Internal Author	12%	88%	6.5 Hours
19.	Chem. Abstracts	100%	0%	N/A
20.	Several	100%	0%	N/A
21.	None	0%	100%	6 Min.
22.	Engineering Index	100%	0%	N/A
23.	Chem. Abstracts	100%	0%	N/A
24.	DDC	10%	90%	14 Min.
25.	Inst. for Sci. Info.	100%	0%	N/A
26.	None	0%	100%	90 Min.
27.	None	0%	100%	90 Min.
28.	None	0%	100%	8 Min.
29.	None	0%	100%	?
30.	None	0%	100%	21 Min.
31.	Several	30%	70%	3 Min.
32.	Diabetes Lit. Index	100%	0%	N/A
33.	DDC, Gov. Printing	60%	40%	13 Min.
34.	None	0%	100%	?
35.	None	0%	100%	N/A
36.	DDC 1473 Form	35%	65%	40 Min.
37.	None	0%	100%	23 Min.
38.	None	0%	100%	N/A, Staff Decision
39.	MEDLARS	100%	0%	N/A

	3. Documents/mo. reviewed for SDI input	4. Percent accepted
1.	5000	100
2.	900	100
3.	1200	75
4.	7200	10
6.	500	100
7.	830	90
8.	25	100
9.	110	?
10.	9000	100
11.	2200	?
12.	330	85
13.	16,000	99
14.	8000 per tape	100
15.	500	100
16.	250	?
17.	500	98
18.	1250	4
19.	500	100
20.	29,000	100
21.	?	96
22.	500	100
23.	500	100
24.	300	15
25.	?	80
26.	5500	97
27.	5500	97
28.	100	90
29	6500	1000
30.	840	80
31.	3000	50
32.	400	100
33.	2500	15
34.	1200	90
35.	500	25
36.	2300	85
37.	400	80
38.	?	?
39.	400	100

Section III. Documents (Cont.)

5.	Vocabulary Free, Controlled (C) and source if C	If Controlled number of terms	Number of terms added	6, 7 Indexing Lower limit	Upper limit	Average
1.	Subscribers N/A					
2.	Subscribers N/A					
3.	C., Int. like EJC	4700	9/1000 doc.	-	-	20
4.	Manual Scanning, N/A					
6.	Subscribes, N/A					
7.	Free					
8.	C., Internal	150	?	-	-	3
9.	C., Ext., 4 thesauri	?		-	-	8
10.	Contractor, N/A					
11.	C., Int. (1200 drugs)	1800	10/wk.	1	10	6
12.	C., Internal	1500	4/wk.	3	48	7
13.	Manual Scanning, N/A					
14.	Subscribes, N/A					
15.	Controlled	3000	?	-	-	15
16.	C., Internal, 1000 idenifiers 2000 subjects	3000	3/wk.	-	116	13
17.	Free			-	200	?
18.	C., Internal	3300	0	1	25	20
19.	Subscribes, N/A					
20.	Free			-	-	titles, abstracts texts
21.	Controlled	2000	?	-	225	?
22.	Subscribes, N/A					
23.	Subscribes, N/A					
24.	C., Internal	?	?	8	14	10
25.	Subscribes, N/A					
26.	C., Internal	18,000	?	-	-	13
27.	C., Internal	18,000	?	-	-	13
28.	C., Internal	15,000	40/wk.	-	-	4.3
29.	C., Internal	15,000	4/wk.	2	15	6
30.	C., Internal + DDC	2500 + DDC	2/wk.	-	-	10
31.	Free					text
32.	Subscribes, N/A					
33.	C., Internal	14,000	7/mo.	3	8	5
34.	C., Internal	70 categories	?	-	-	N/A
35.	C., Internal	3000	?	-	-	4 + modifiers
36.	C., Internal	13,500	150/yr.	1	100	14
37.	C., Int. like AICHE	8000	40/mo.	-	-	24 (stored of 40)
38.	Manual Scanning, N/A					
39.	C., Internal	7000	50/mo.	-	-	30

	8. Number of one-time articles per month		9. Number of journals per month		10. Use Contractors
	Internal	External	Internal	External	
1.	N/A				
2.	N/A				
3.	1200	0	0	0	no
4.	0	360	0	6840	no
6.	N/A				
7.	0	822	0	8	no
8.	25	0	0	0	no
9.	0	11	1	98	yes (pilot test)
10.	N/A				yes
11.	0	0	0	2200	no
12.	8	157	16	149	yes
13.	0	16,000	0	0	no
14.	N/A				
15.	0	0	0	500	yes
16.	250	0	0	0	no
17.	0	0	0	500	no
18.	26	8	0	19	no
19.	N/A				
20.	0	0	0	30,000	no
21.				all	no
22.	N/A				
23.	N/A				
24.	60	15	0	225	no
25.	N/A				
26.	1417	4083	0	0	yes
27.	1417	4083	0	0	yes
28.	0	0	0	100	no
29.	?	?	?	?	yes
30.	168	479	0	193	no
31.	1446	1446	15	92	no
32.	N/A				
33.	0	1750	0	750	no
34.	0	0	5	1195	no
35.	80	420	0	0	no
36.	345	1955	0	0	no
37.	400	0	0	0	no
38.	some	some		most	no
39.	260	140	0	0	no

Section IV. Profiles

1. Profile vocabulary construction	2. Number of Profiles maximum allowed	In System	3. Limit of terms Lower	Upper	Averag
1. DDC vocal.	1	4000	-	24	
2. NASA/SCAN list	N/A	Serves 200		189 topics	
3. Internal vocab.	usually 1	10	-	-	
4. Free language	only 1 necessary	12	-	-	
6. Free language	open	30	-	-	
7. Free, staff aid	open	114	-	-	
8. 3 int. Thesauri	1	800	-	-	
9. 1 ext. thesauri	5	30	-	-	
10. 1. Doc. Inc. vocab.	open	?	1	150	49
2. Subject list	N/A	?	-	-	
11. Free	1	?	-	-	
12. Internal vocab.	3	1500	-	-	
13. Free	1	159	-	-	
14. MESH (medical)	1	38	-	400	
15. Controlled categories and terms	experimental	?	-	7 diff. phrases	
16. Free, translated by staff	only 1 necessary	260	-	1000	
17. Free, staff aid	many encouraged, open	386			rarely ceeds
18. Internal, staff aid	1	2000	1	20	
19. Free, staff aid	open	240	-	-	
20. Free	99	220	-	9999	130
21. Free	36	60	1	60	
22. Engineering index	100	?	-	-	
23. Free	20	42	-	-	
24. Internal vocab.	6	300	-	-	
25. Free, staff aid	open	250	1	1000	
26. Internal vocab.	usually 1	850	-	-	
27. Int., NASA/SCAN list	N/A	serves 500	-	189 categories	
28. Uses research grant	1	103	-	-	18
29. Internal	usually 1	218	-	-	
30. Internal DDC	open	195	-	300	
31. Free staff aid	1	2800	-	-	
32. MESH staff aid	open	11	1	200	80.4
33. Internal, staff aid	open	700	1	8	4
34. Internal IBM/SCAN	N/A	?	-	70	
35. Internal, staff aid	2 if necessary	225	-	-	5
36. Internal, staff aid	open	446	-	-	13
37. Internal, staff aid	open	**	5	range 265	125

38. Free	only 1	necessary	?	-	-	
39. Internal, staff aid	1	100		-	-	10

* In addition user submits 1 page job and duties summary that staff index and add to profile

** Ready to run, but non-operational. If system were to go on-line, 200 profiles would be used.

	4. Use Weights or logic	5. Can user get automatic distrib. of named items?	6. Does user suggest document sources?
1.	neg. *	no	no
2.	no	no	no
3.	links, roles	yes	yes
4.	no, unnecessary, manual	yes	yes
6.	no	no (done outside system)	no
7.	and, or, not	no (done outside system)	yes
8.	hierarchal alpha code	yes	yes
9.	0 to 0.999 (0 is neg.)	yes	yes
10.	(1) -9 to +9 (2) no	no	no
11.	no	yes	yes
12.	(1) scale options (2) and, or, not	yes	yes
13.	no, unnecessary, manual	yes	yes
14.	boolean threshold 10 exactly	no	no
15.	0, 1, 2 for descriptors frequency count for categories	yes	no
16.	and, or, not, hierarchal, probablistic or	yes	no
17.	1 to 10, boolean operators, root matching	yes	yes
18.	max. 4 "s	no	no
19.	and, or, not; option 1 to 1000	yes	no
20.	0.0001 to 0.9999 & neg. root matching	yes	yes
21.	-9 to +9	yes	yes
22.	no	yes	yes
23.	no	yes	yes
24.	1 to 9 and -1 and root matching	yes	no
25.	no	yes	no
26.	weights simulating, boolean logic	yes	yes
27.	weights simulating boolean logic	yes	yes
28.	hierarchal decimal code	no	no
29.	no	yes	no
30.	no	yes	no
31.	no	yes	yes
32.	no	yes	no
33.	no	no	no
34.	no	no	yes
35.	uses modifiers as screen	no	yes
36.	yes, varies	no	yes internal
37.	-9 to +9, 0 = no significance	yes	yes
38.	unnecessary, manual	?	?
39.	and, or, not	yes	yes

Section V. Profile Modification

1. Method of modification	2. How often	3. Adding or deleting terms (A), modifying weights (M), Both (B)	4. How profiles are modified
Print out of annotated profile	6 mo's.	B	user
User request	each run	A	user
Monthly interview	batched to 3 mos.	A	
User request	as necessary	A	
Interviews	3 mo's.	A	
Interviews based on record of response	each run	A	
Returnable notice	each run	A	
Returnable notice and interview each run	each run	B	staff
1. Custom prof.-returnable	1. each run	1. B	1. computer aids
2. Subject list-interview	2. 3 mo's.	2. A	2. user picks documents
Interviews based on responses	monthly	A	
Returnable notice	each run	B	system
Send profile	3 mo's.	A	
Interview based on poor results	each run	B	staff & user
Interview or request	6 mo's.	B	system with user feedback
Returnable & staff initiative	each run	B	staff
User request	each run	B	staff & user, scale and logic
Relevance measure	not often	B	staff evaluation or request
Interviews	every 2 wks.	B	staff
Returnable notice	3 times/wk.	A	
Returnable & interview	daily & mo.	A	
Returnable notice	each run	A	
Interviews	each run	A	
Returnable notice	each run	B	system thru user responses
Returnable & interview	each run & yearly	A	
Interview - based on irrelevant citations	each run	B	staff & user
User request	each run	B	rarely necessary
Returnable notice	each run	A	staff checks responses & makes changes if possible

	1. (Continued)	2. (Continued)	3. (continued)	4. (Continued)
29.	Request & interview	each run & 3 mo's.	A	
30.	Returnable & year.	each run	A	
31.	Returnable notice	each run	A	statistical sheet
32.	Returnable notice	each run	B	
33.	Interviews	as necessary	A	
34.	Returnable & interview	rarely necessary	A	
35.	User request & interview	each run & 18 mo's.	B	
36.	Returnable notice	each run	B	user or staff & list of doc. terms
37.	Returnable notice	each run	B	change role indicators
38.	User request & questionnaire	each run & 6 mo's.	A	
39.	User request & interview	each run & 6 mo's.	B	staff

	1. Method of matching A, B, C, D*	2. Threshold			3. Comments
		Does it change?	If Yes Controlled by user or system	varies user or doc.	
1.	D	no			Neg. user term scratches match doc* & pos. user te
2.	DONE BY NASA				
3.	A	no			1 hit; links & roles
4.	JUDGMENT				specialist understands clients' needs & scans doc
6.	A	yes	system	user	no weights
7.	A	no			boolean logic; 1 hit
8.	A	no			1 hit
9.	A	no	user	user	user assigns terms & hit level on 1-9 scale; altered to level 0.999
		no for list			
10.	A	yes	system	user	according to user request
11.	A	yes	user	user	user supplies instructions
12.	A	1-yes 2-no	system	user	1-target score 2-boolean logic 1 hit
13.	JUDGMENT				mostly engineering drawin scanner determines relev
14.	A	no			boolean threshold of 10 ex
15.	A	no			1 match
16.	A	yes	staff	user	extended boolean expressi
17.	A	yes	system	user	weights & boolean operato determine hit level
18.	D	no			3 term or 1*
19.	DONE BY CHEMICAL ABSTRACTS SERVICE				
20.	A	yes	system	doc.	depth of doc. input determ hit level

23.	A		user	user	hit between 1 & 9 (scale -9 to -9)
24.	A	no			boolean logic
25.	A	no			boolean logic
26.	C	yes	system	user	matched terms ÷ doc. terms needs to be equal or greater than hit level
25.	A	no			1 hit
26.	A	yes	analyst (system)	user	groups within each profile have indep. hit levels
27.	A	no			one linear match
28.	B	no			
29.	D	yes	user	user	
30.	A	yes	both	user	% of doc. terms match profile
31.	A	yes	user	user	
32.	A	no			
33.	A	no			
34.	C	no			altho the center is mechanized, SDI is manual
35.	J U D G M E N T				library reviews quality & quantity with user's OK
36.	A	yes	either	user	set low to start & adjusted until user is satisfied
37.	D	yes	system	user	senior librarian scans all incoming doc's.
38.	J U D G M E N T				non-computerized equipment is used & staff supply final judgment.
39.	C	no	user	user	

* A = independent user and document profiles, B = independent user profiles and aggregated document profiles; C = aggregated user profiles and individual document profiles; D = aggregated user and document profiles

1. Content	2. Form	3. How often Notices are Sent
1,2,6,7,8,DDC number	1, full memo	2X/mo.
1,6,7, complete descriptive catalog	1	2X/mo.
1	1	2X/mo.
A C C O R D I N G T O C L I E N T		weekly or monthly
1,2,3	1 clipped to abstract issue	weekly
1,2,3,7 profile print-out	1	every 2 weeks
5,6	1,5 or cover sheet	daily
1,2,7,8	3 with xerox abstract	4 runs in 9 mo's.
1,2 for both; 7,8 for tailored profile	?	2X/mo.
1,2,7	1, [2 to division heads]	daily, weekly
1,2,3,6,7	1,2	every 2 weeks
8	1,5	daily
1,4,5,7	2	every 2 mo's.
1,2 feedback [method not determined yet]	1 in rank order, 5	every 14 days
1 if classified 2,3,7,8,9	1	monthly
1,2	1,3 double perforated	weekly
1,2,3,5,6,7,8	3	monthly
1,3	1	2X/mo.
3,4,5 [max 9],6,7, first 1040 input characters	1,3 double perforated	weekly
1,2,3,4,5,6,7	2,3	daily
1,2,3,5,6,7,8	1, all abstracts for comment	monthly
1,2,3	1	every 2 weeks
1,3,4,5,7,8	1,3	monthly
1,5,8	1,2	weekly
1,3,6, [or notation of contents],7	1	weekly
1,3,6,7	1	2X/mo.
first page with 1,2,4,5, translation	first page	every 2 weeks
1,2,3,5,6,8	1,5	weekly
1,2,3,7	1,2,3	weekly
1,2,3,7,8	1,2,3	weekly
1,3,4,6,7	1	monthly
2,3	1	monthly
1,3,6,8, distribution card	1	weekly
0	5	daily
1,4,6,7	4	2X/mo.
2 [not printed], 3,4,5	1 on paper, abstract typed	2X/mo.
0	5 or abstract if no doc.	daily
2	?	monthly

1. Anderson, R. R., A Profile Guide For Ames LAB SDI Users, Ames Laboratory, Undated.

2. Armstrong, D. L. and M. T. Grenier, A Central Information Retrieval System, Journal of Chemical Documentation, 5, 99 [1965].

3. Balz, Charles F. and Richard H. Stanwood, Merge: A Current-Awareness and Retrospective Searching System for Technical Documents, International Business Machines Corporation, Owego, New York, August 21, 1963.

4. Bivona, William A. and Edward J. Goldblum, Selective Dissemination of Information: Review of Selected Systems and A Design for Army Technical Libraries, Information Dynamics Corporation, Reading, Massachusetts, August 1966.

5. Buchanan, J. R. and F. C. Hutton, Analysis and Automated Handling of Technical Information at the Nuclear Safety Information Center, U. S. Atomic Energy Commission, Oak Ridge, Tennessee, June 1966.

6. Buchanan, J. R. and F. C. Hutton, Analysis and Automated Handling of Technical Information at the NSIC, Oak Ridge, Tennessee, Atomic Energy Commission, Winter 1966 - 1967.

7. Croxton, F. E., Automation Progress at RSIC - The Status of Alpha I, U. S. Army Missile Command, Red Stone Arsenal Alabama, 2 October 1963.

8. Ebersole, J. L. and J. M. Worth, North America Aviation's National Operating System, North American Aviation, Inc., Undated.

9. Gallagher, Joan L., Experience leading to the Introduction of an SDI System, The Journal of Chemical Documentation, Vol. 6, No. 3, pp. 151-153, August 1966.

10. Gatto, James J., Application of Computer-Based Retrieval Concepts To a Marketing Information Dissemination System, Westinghouse Electric Corporation, East Pittsburg, Pennsylvania, Proceedings - Spring Joint Computer Conference, 1966.

11. Goldwyn, A. J. and G. Ember, DIAPAS: A Personalized Alerting Service for Diabetes, Reprint Methods of Information In Medicine, Vol. 6, No. 3, pp. 130-135, July 1967.

12. Henderson, Paul B. Jr., A Computer-Generated Thesaurus, Westinghouse Electric Corporation, Pittsburgh, Pennsylvania, May 27-29, 1964.

13. Henderson, Dr. Paul B. Jr., An Operational Language for User-Oriented Expression of Data Retrieval Specifications, Westinghouse Electric Corporation, Pittsburgh, Pennsylvania, September 2-7, 1966.

14. Henderson, Paul B. Jr., Management's Role in Developing and Using Computerized Personnel Information Systems, Westinghouse Electric Corporation, Pittsburgh, Pennsylvania, February 3-5, 1965.

15. Henderson, Paul B. Jr., On the Design of Data Systems for File Analysis and Information Retrieval, Westinghouse Electric Corporation, Pittsburgh, Pennsylvania, March 12-14, 1964.

16. Hifi, Jack W., Matching of Descriptors In A Selective Dissemination System, Office of the Chief Engineer, Bureau of Reclamation, Denver, Colorado. October, 1963.

17. Horvath, Paul J., Robert F. Carroll, Joyce Dlugos, and Alice Yanosko, The Unified Mechanization of a Thesaurus, A Dual-Dictionary, and A Selective Dissemination System, B. F. Goodrich Chemical Company, Computer Department, Cleveland, Ohio, April 1966.

18. Hoshovsky, A. G. and C. S. Downie, Selective Dissemination of Information In Practice - Survey of Operational and Experimental SDI Systems, Office of Aerospace Research, 21 September 1967.

19. Hunter, C. H. and W. S. Stone, TIGIR, The TIG Information Retrieval Program for the Philco 2000, Knolls Atomic Power Laboratory, July 15, 1963.

20. Kidder, Major Arthur, Selective Dissemination of Information, Air Force Systems Command, Andrews Air Force Base, 1965.

21. Levy, N. P., An Experimental Study to Evaluate Protype General and Selective Dissemination of Information Systems, Western Electric Company, 1963.

22. Luhn, H. P., Selective Dissemination of New Scientific Information with the Aid of Electronic Processing Equipment, AM Document, 12 [2] 131-38 [1961].

23. Peterson, Norman D., Information Dissemination and Retrieval at Bonneville Power Administration, Bonneville Power Administration, Portland, Oregon, Undated.

24. Rees, Alan M., The Evaluation of Retrieval Systems, Western Reserve University, Cleveland, Ohio, June 14-17, 1965.

25. Rees, Alan M., The Measurability of Relevance, Western Reserve University, Cleveland, Ohio, October 3-7, 1966.

26. Rees, Alan M., The Relevance of Relevance to the Testing and Evaluation of Document Retrieval Systems, Western Reserve University, September 13, 1966.

27.. Rees, Alan M. and Douglas G. Schultz, A Field Experimental Approach to the Study of Relevance Assessments in Relation to Document Searching, Western Reserve University, Undated.

28. Resnick, A. and C. B. Hensley, The Use of Diary and Interview Techniques in Evaluating A System for Disseminating Technical Information, American Documentation, Vol. 14, No. 2, pp. 109-116, April 1963.

29. Sage, C. R., Comprehensive Dissemination of Current Literature, Institure for Atomic Research, Iowa State University, Ames, Iowa, American Documentation, - Vol. 17, No. 4, p. 155-177, October 1966.

30. Sage, C. R., The Utilization of a National or Regional Current Awareness System for Retrospective Information Retrieval, Institute for Atomic Research, Mathematics and Computer, Iowa State University, Ames, Iowa, Undated.

31. Sage, C. R. and R. R. Anderson, Ames Laboratory Selective Dissemination of Information Computer System - A Survey of User Evaluation, Ames Laboratory of the U. S. Atomic Energy Commission, Undated.

32. Sage, C. R. and R. R. Anderson, and C. H. Cameron, Ames Laboratory SDI Systems Reference Manual, Ames Laboratory, U.S.A.E.C., August 1964.

33. Sage, C. R., R. R. Anderson, and D. R. Fitzwater, Adaptive Information Dissemination, American Documentation, Vol. 16, No.3, p. 185-200, July 1965.

34. Sage, Charles and D. R. Fitzwater, Ames Laboratory Selective Dissemination of Information Computer System, Ames Laboratory, September 1963.

35. Saracevic, Tefko and Alan M. Rees, Towards the Identification and Control of Variables in Information Retrieval Experimentation, Western Reserve University, October 10-15 1965.

36. Schmidt, C. J. and R. G. Murdick, Setting Up A Technical Information System, Knolls Atomic Power Laboratory, General Electric Company, Schenactady, New York, May 10, 1963.

37. Schmidt, C. J., R. P. Wesley, and C. H. Hunter, Kapl's Automated Information Retrieval System, Knolls Atomic Power Laboratory, Paper presented at the 1964 American Documentation Institute annual meeting.

38. Schneider, John H., Proposed Categories for Classifying the Subject Content of Research Grants on a PHS-Wide Basis, National Institutes of Health, August 1964.

39. Segarra, Carlos O., An Approach to Cost Effectiveness Of A Selective Mechanized Document Processing System, U. S. Army Mobility Equipment Command, Fort Belvoir, Virginia, AD 651486, March 1967.

40. Sprague, Ralph H. Jr., A Comparison Of Systems For Selectively Disseminating Information, Foundation for Economic and Business Studies, Indiana University, June 1965.

41. Young, E. J. and A. S. Williams, Historical Development and Present Status - Douglas Aircraft Company Computerized Library Program, Douglas Missile and Space Systems Division, May 1965.

42. _____, An Experiment in Selective Dissemination of Information, Chemical Abstracts Service of the American Society, Columbus, Ohio, July 11, 1966.

43. _____, CBAC Tapes and Search Services, Chemical Abstract Service, Columbus, Ohio, April 6, 1966.

44. _____, CT Tapes and Search Services, Chemical Abstracts Service, Columbus, Ohio, September 1966.

45. _____, MERCURY System for Document Dissemination, Bell Telephone Laboratories Technical Information Libraries, August 11, 1967.

46. _____, POST Tapes and Search Services, Chemical Abstracts Service, Columbus, Ohio, September 1966.

47. _____, Program Documentation For A Selective Dissemination System For NASA Scientific and Technical Information, NASA, CR-62021, June 1966.

48. _____, Scientific and Technical Information Division Technical Document Center, U. S. Army Engineer Research and Development Laboratories, Fort Belvoir, Virginia, Undated.

49. _____, SDI Profiling Handbook, ECOM, Fort Monmouth, New Jersey, undated.

50. _____, Selective Dissemination of Information, IBM, 1962.

51. _____, Selective Dissemination of Information [SDI], U. S. Army Electronics Command, Fort Monmouth, New Jersey, undated.

52. _____, Selective Dissemination of Information [SDI], U. S. Army Electronics Command Technical Information Division, Information Office, Fort Monmouth, New Jersey, 1 October 1966.

53. _____, Selective Dissemination of Information [SDI], Volume I, U. S. Army Natick Laboratories, Natick, Massachusetts, May 1967.

54. _____, Selective Dissemination of Information [SDI], Volume II, U. S. Army Natick Laboratories, Natick, Massachusetts, June 1967.

55. _____, Standard Interest Profiles, Aerospace Research Applications Center, Undated.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Office of Aerospace Research Information Studies Division Office of Scientific and Technical Information Arlington, Virginia 22209	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

SELECTIVE DISSEMINATION OF INFORMATION [SDI] Analysis of Experimental and Operational SDI Services, 1967

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific - Final

5. AUTHOR(S) (First name, middle initial, last name)

Alexander G. Hoshovsky

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
20 July 1969	69	55

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) OAR 69-0016
d.	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
Tech-other	Office of Aerospace Research 1400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

The objective of this study is to learn more about the options available for the application of the Selective Dissemination of Information [SDI] concept prior to decisions about its possible use in Air Force research and development activities. The data collection and verification tools consist of structured questionnaires, phone interviews, and published literature on specific SDI systems. The findings and conclusions of this report have been drawn from forty-one of some fifty known SDI systems in operation as of August 1967, and therefore represent a fair sample of the total population.

The findings support the contention that SDI concepts and methodologies can be made to work for any number of users with any volume of document input; however, none of the systems are even close to their capacity--only a small portion of eligible users depend on SDI services.

All surveyed systems managers believe that SDI serves a real and important need of screening the literature for specialized users and they intend to continue some form of current awareness. The SDI concept has been successfully applied to primary distribution of laboratory reports and is likely to effect the future pattern of technical journal production and distribution.

DD FORM 1 NOV 1473

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Selective Dissemination of Information [SDI]						
Document-altering service						
Profile matching techniques						
User notices						
User populations						
Controlled vocabulary						
Free vocabulary						