RANDOLPH E. HOCK

Providing Access to Externally Available Bibliographic Data Bases in an Academic Library

The ready availability of externally processed bibliographic data bases has made it possible for an academic library to provide computerized searches on a large number of data bases with a very small initial investment and utilizing its own personnel. The experience of the University of Pennsylvania Libraries has confirmed that such an approach is indeed feasible. This article discusses the approach, questions and problems encountered, and the factors considered in their resolution. Also discussed are the role of the data services librarian, the costs incurred, and some observations as to the philosophy of the approach, with particular attention to the integration of the service into the reference department.

THE INCREASING AVAILABILITY of large machine-readable data bases has for several years served as an indicator to libraries that a means of greater information retrieval services to their users was forthcoming. The role of libraries in providing this type of service was discussed in a recent article by Richard DeGennaro.[1] In that article he wrote about the possibilities of the library's serving as a "broker" rather than as a processor for data bases. He pointed out that libraries will play a key role in providing access to data bases, but that the in-house processing approach is not feasible due to the high cost involved and the nature of the demand. Also stressed was the likely trend toward the utilization of a mixture of batch-mode and on-line services, each serving specific functions.

In the past two years the University of Pennsylvania Libraries, as well as some other academic libraries, have been pursuing this approach. The general goal has been to provide computerized literature searches for the university community, with a minimum investment, by taking advantage of currently available external services.

In this article an attempt will be made to summarize the situation faced by academic libraries undertaking such an approach. The emphasis will be on the decisions which are necessary, the costs involved, and some general observations which should be of help to those contemplating accessing such systems. Technical details regarding the systems themselves will be avoided, since

At the time of the preparation of this article, Randolph E. Hock was data services librarian, University of Pennsylvania, Philadelphia. He is now pursuing further graduate study.

such information is available elsewhere. An excellent up-to-date source of detailed information about on-line systems is the recent book by Lancaster and Fayen.[2]

At the outset of projects of this type, certain policy decisions become necessary which are potentially crucial to the success of the endeavor. Among the first that arises is the question of where within the library structure such a service should be located.

The Service as a Reference Function

After consideration of the various alternatives (organizationally and physically) within the library system, the conclusion has been reached that the reference department is the most reasonable location. There are two primary reasons for this. First, providing bibliographic services *is* a reference function; and, second, in order to give users fullest advantage of total library services this office should not be separate from, or in competition with, other reference functions of the library. To further this "integrated" approach, it has been found feasible and desirable to involve the entire reference staff, particularly by delegating responsibilities for the searching among the staff, each specializing in one or more data bases.

Alternatively, the service could be located within the library's systems office, the rationale for this approach being the computerized origin of the services. However, for the reasons given above, the fact that handling these services requires no systems experience, and the fact that the systems office is usually remote from the public services division of the library, the former approach is preferred.

Types of Systems Available

The prospective "broker" of externally processed data base services is confronted with three main categories of service which can be used either separately or in combination. These are batch-mode (off-line) services, on-line bibliographic services, and on-line nonbibliographic ("statistical") services.

The batch-mode services are the easiest (for the librarian) to use. Providing access to a large number of these data bases can be as uncomplicated as collecting data base directories[3–6] and then providing this information to users as the occasion arises.

The on-line data bases require more effort to access, considerably more participation on the part of the librarians, and a number of other considerations which will be discussed below. However, the actual amount of service which can be provided to the user will not only be greater, but is more likely to be utilized. The second type of on-line service, statistical data bases (such as Predicasts, the National Bureau of Economic Research Time Series Data Bank, and STOCKRETRIEVE), presents more complicated problems, not in terms of the actual access, but in the on-line manipulation required due to the contents of the data base. The advantages are that the material retrieved is the actual information the user is looking for, rather than bibliographic references. This specificity of information has an additional drawback, however, in that each data base is likely to be of interest to but a small portion of the user community.

Utilizing Batch Services

Following considerations of the above points, the next step is the actual implementation of services. The discussion which follows will attempt to generalize where possible, but will also rely on presentation of actual situations encountered and experiences gained. As indicated, taking advantage of batch services involves, first, the collection of information on those available. In addition, the library may also become involved in promoting the services and as-

sisting users in choosing the appropriate data base, making the contacts, filling out the request forms, and evaluating results.

Unfortunately, none of the data base directories, though helpful, contain all the information which is frequently called for in fulfilling the above functions. They do, nevertheless, serve as an adequate source for obtaining further information and price lists.

Following this approach, information was compiled on those data bases which were readily available, and which were likely to be of interest to a number of potential users (by "available" it is meant that they could be accessed in already processed form simply by contacting a center either by mail or by phone). No attempt was made to produce a total list of either all data bases or all processing centers. It was found that twenty-nine available data bases might be of use and that these were available either directly from the data base producers or from one or more of six processing centers.

This information was gathered, and considerable time was spent in the promotion of the services. Working at this point only with the batch services, the promotional emphasis had to be placed on selective dissemination of information (SDI) services rather than the retrospective searches since the prices for SDI ($40/year and up) were much more appealing than the $200 which was typical for a four-year retrospective search. After a few months, only a handful of people had subscribed to such searches. The reasons given for not subscribing by most people contacted were that funds were not available and that they were satisfied with their present information gathering methods. Another reason, which seemed obvious, though not stated directly, was that SDI services have yet to prove their value to university people.

Selection of On-Line Systems

Besides the limited demand for those services which can be provided in batch mode, several factors indicate the advisability of utilizing on-line services in addition to the batch services, in spite of the greater degree of involvement demanded of the library. These factors are: (1) the cost of searches: on-line retrospective searches (in most cases) are significantly less expensive in comparison to the batch searches; (2) the turnaround time: a few days or immediately for on-line, a few weeks for batch; (3) the interactive capability: only available on-line; and (4) the rapid increase in the number of on-line data bases.

More specifically in regard to the price differential, to get a typical retrospective search done on *Chemical Abstracts* for 1970–1973 would cost $640 from one leading center and $400 from another. The same search could be done for from $5 to $40 through the on-line services. A typical retrospective search for the NTIS files would cost $50 from NTIS directly, $65 from one center, and $50 from another, but less than $15 through the on-line services. (Figures for the batch services are based on the promotional literature supplied by the centers. Figures for on-line costs are based on searches actually done.)

Once the choice has been made to implement on-line services, the next decision necessary is which system(s) to access, and which of their data bases to use. At the present time there are three on-line bibliographic literature searching systems which are actively promoting their services. These are Lockheed Information Systems, New York Times Information Bank, and Systems Development Corporation (SDC).

Lockheed and SDC stand out immediately as likely choices because of the data bases they offer, the range of sub-

ject areas they cover, and the lack of any subscription charge or minimum fee. Based on the areas of study of the graduate students currently enrolled at this university, it was found that one or more of the data bases offered by those two systems could be of potential use to approximately 48 percent of that group.

Likewise, the potential user group for the New York Times Information Bank is quite large, and this data base is consequently quite appealing. The major difference, though, in accessing this system as compared to the other two is the minimum investment required. Regardless of how little the system may be used, a minimum monthly cost of something in excess of $300 is incurred.

Due to this necessary financial commitment for the New York Times, utilization of that service by this library was delayed until some experience was gained with the other two systems. However, within a few months after the implementation of the first two systems, the New York Times was also accessed. The decision was precipitated by the announcement of a considerable decrease in minimum charges.

Pricing Structure

The primary factor to consider when deciding what to charge users for searches is how much the library is willing to spend of its own money. If the amount is equal to what would be spent in developing one's own information retrieval system, the library can probably give away hundreds of searches per year absolutely free. Disregarding that option, the library can decide to offer free searches and limit demand in any one of several ways. This approach serves the very useful function of stimulating use of the service. However, response to anything free is usually not a good indicator of either potential demand or actual need. Getting some feel for these two factors is of considerable importance at this stage in the development of these services.

At the opposite extreme, a library can decide to recover all costs including connect time, communications, personnel, space, etc. This approach has the disadvantage of putting the cost of a search out of the reach of the graduate students who perhaps need the service more than any other group. Also, if total cost recovery is a valid rationale for this service, is it also a valid rationale for all library functions?

Taking the approach that the searches should be partially subsidized, a reasonable pricing structure, which has proved workable, is, first, to offer the services at a flat rate for a few months, and then, at the end of that period, when a nucleus of subscribers has been created, to change the pricing structure to reflect costs more accurately. In both stages a criterion should be that the price schedule should be easy for the potential user to understand. This approach was tried, using $10 per search as the flat rate (a "search" being defined as coverage of one data base with output limited to 50 citations with abstracts, or 100 citations without abstracts). Additional output was offered at $.10 and $.05 respectively per citation. This price structure covered the direct costs (connect time and communications) of some searches and fell far short for other searches. However, it allowed the expenditure of the approximate amount of "seed money" anticipated in order to advertise the service.

This price schedule was maintained through one term, at which time the structure was changed to give the user two options, designated as "standard searches," and "special searches." With the standard search, the user is charged a flat rate and is allowed to choose up to a certain number of descriptors (the rate and number of descriptors varying

with the data base, but typically being $10 and ten descriptors). Output of up to fifty citations is included in the base price. For these searches the user is not present. For the special searches the user is present and is charged the actual cost of connect time and printing. As with the standard searches, the search strategy and input terms are worked out ahead of time, but are modified to whatever extent the user desires as the search progresses.

Promotion and Procedures

Since the type of service being offered is one with which members of the academic community are generally not familiar, but one for which the need should at least be explored, it is felt that a fairly vigorous promotional program should be undertaken. This program can involve the campus news media, flyers, posters, displays, direct personal contact, contact through departmental librarians, and word-of-mouth referrals among users. Expenses for all of these approaches can easily be limited to personnel time and the library's own duplicating services.

One other approach which is being tried is to make the grant applicants aware of the services at the time they are preparing their budgets. To accomplish this, an appendix describing the service was prepared for the handbook distributed by the university office which handles grants. The appendix describes the retrospective searches, the SDI searches, and also the "Research in Progress" searches which are available through the Smithsonian Science Information Exchange.

In developing procedures, the question as to what extent to utilize user participation has been one of the most difficult determinations to make. The extremes would have been either a totally delegated search, in which the requestor is not present, or a totally nondelegated search, in which the intermediate (the library staff member) need not be present. The obvious disadvantage to the delegated search is that both recall and relevance can suffer. The advantage is that the search can be done in less time (with resulting decrease in cost). The disadvantages to the nondelegated search are the necessity for training every requestor in use of the service (at great cost), and the likelihood that many, if not most, occasional users would not be able to take full advantage of the systems capabilities (with a resulting decrease in recall). The advantage is an ultimately better performance for the frequent user (if any individual is willing to spend the money and time to become really familiar with the system). One solution is to take an intermediate approach and have the reference staff serve as "middlepersons," but to try to have the user present. This would result in lowest training costs, the availability of an experienced searcher for every search, and the availability of the requestor to give feedback as the search progresses. The impression has been gained that requestors who were actually present for the search seem frequently more satisfied than those who opted to delegate the search. This degree of satisfaction seems to be due not only to the increased relevance of the retrieved set but also to the fact that the requestor has a better understanding of the capabilities and limitations of the system. The usefulness of the "middleperson" approach has also been indicated by Borman and Mittman.[7]

The policy of encouraging the requestor to be present for each search has been found to have several shortcomings: (1) some users find it inconvenient to be present; (2) the total turnaround time is usually greater due to the necessity of waiting for a time convenient for both the searcher and the requestor; (3) some searches are so simple that the requestor's presence is unnecessary; and (4) connect times are

usually much greater with the requestor present, with the consequent higher cost of searching. It is recognized, however, that for many searches direct user participation is still essential for most efficient retrieval. This conflict can be resolved by offering the two options described in the section on pricing. The standard search (flat rate, user not present) permits availability of quick, inexpensive searches, especially for students, and the special searches permit more thorough searching utilizing more fully the total system capabilities.

Another procedural decision is how many searchers to use. In most cases the volume of searches will be small enough to require only one or two searchers, but on the other hand, there are two main advantages to distributing the responsibility among most or all of the reference staff. First, it helps to keep the service more integrally connected with the reference department; and, second, variation in content and indexing of the data bases is so great that the more data bases an individual attempts to handle, the less intimately each can be known, and the less efficient the retrieval process. This approach has been met with enthusiasm by reference staff members, even though most have had no prior experience with computers.

Functions of the Data Services Librarian

The function of this person (called information services librarian in some institutions) can be broken down into five areas: coordination, promotion, training, searching, and consultation.

Coordination includes development of procedures, liaison with and among searchers, and liaison with departmental librarians and other related services on campus.

Promotion is the one function which is probably most alien to many librarians. Nevertheless, if it is indeed felt that the services being offered are time saving and worthwhile, one should either be willing to put considerable effort into their promotion, or rethink the premise. Over the next few decades, patterns of scientific communication are sure to change in this direction, but if these services are not exploited at present, many valuable resources are going to be wasted in the interim.

Training in the use of the on-line systems consists essentially of digesting the manuals and accumulating experience. Once one person within an institution has become familiar with the systems, that person can train others easily. Training courses are offered by both Lockheed and SDC, and experience has shown these to be helpful; but if one is capable of reading the manuals (which assume only a rudimentary knowledge of systems terminology), one can proceed from there and, within a couple of hours of connect time, be well on the way to proficiency. Again, the knowledge of the operation of the systems is but one part of what is to be learned, the other main part being knowledge of the indexing and contents of the individual data bases. Manuals for both SDC and Lockheed could be of greater help if they included details about the indexing policies followed (or allegedly followed) by the data base producers, and also some details about the posting procedures followed in putting those data bases onto the on-line indexes.

It was found that an average of 3.3 hours of connect time was necessary per searcher in preparation for his or her first actual search. Additional training time was accumulated in the first few searches and is reflected in the difference in average search time for the first month and the second. The average time decreased by eleven minutes during that period. Charging this difference to training costs gives a total training (connect) time of less than four hours per person, or $140–$280 of connect

time (depending on the data bases used).

The consultation function served by the data services librarian tends to serve simultaneously as an in-depth reference function. Primarily, the consultation is to provide potential users with information about the computerized searches, but this information is more valuable to the user if supplemented by further suggestions as to the approach to his or her literature problem. Such personal attention can be had at the reference desk, but is frequently qualitatively and quantitatively insufficient due to lack of time.

Providing the information about the computerized searches requires that the data services librarian be familiar with what services are available, what the output looks like, the turnaround times, which data elements are included, the types of literature covered, and specifically which journals are covered. For the batch services this information can be gotten mainly from the promotional literature of the centers. For the online services, this information is available from the manuals, the promotional literature, and the hardcopy of the data bases.

Costs

A cause for considerable apprehension when approaching these services is how to anticipate real costs correctly. Although the systems sound relatively inexpensive, few figures are available except from the suppliers. Fortunately, there are no large cost categories which are not readily anticipated. Costs can be controlled to any degree a particular library desires, depending upon how quickly the library wishes to offer the services, how many of the data bases are offered, how many searchers are trained, the sophistication of the terminal used, and the degree to which the library is willing to subsidize the services.

Fixed costs include equipment rental (or purchase), service contracts for equipment, training time, and overhead. Equipment can be rented for about $80 per month on up, and service contracts are about $20 per month. Overhead is the same as for other library operations. Personnel costs will vary, depending upon how quickly and how deeply the library wishes to become involved.

The variable costs depend upon how many searches are done and how long the searches take. For a typical search, about thirty minutes of the coordinator's time is required (talking with the requestor, handling printout, billing, etc.). A search typically requires from ten to thirty minutes for its execution (including discussion, logging on, and connect time). Printing costs vary from $.05 per page up to $.25 per citation. Typical searches retrieve between ten and sixty citations. It has been found that the average direct cost-per-search on the ERIC data base, which is one of the least expensive, but also the most popular, is less than $12. Searches on the other data bases are greater, approximately in proportion to the cost of connect time.

Summary

Experience with accessing externally available bibliographic literature searching services indicates that such an approach is feasible (perhaps even obligatory) in the academic library, without waiting for further developments or extensive studies. Costs are reasonable, and perhaps of more importance, subject to whatever controls may be applied. User acceptance is good, though not overwhelming, and the approach of supplying "seed money" seems to be effective since an increasing proportion of users have been referred to the service by former users.

As to how these services fit in with users' needs, it is felt that many undergraduates can use the services as a basic source for some of their papers. For

graduate students, faculty, and other researchers, the service can fulfill an important function as a supplement to their usual information retrieval methods (as indicated by Back[8]).

As more experience is gained, as more formal studies are completed, and as suppliers become more aware of the academic market, the ease and efficiency with which the services can be provided will increase. The present state of the art is such, however, that any academic library with sufficient interest can successfully serve as a "broker" for these services.

References

1. Richard DeGennaro, "Providing Bibliographic Services from Machine-Readable Data Bases—The Library's Role," *Journal of Library Automation* 6:215–22 (Dec. 1973).
2. F. W. Lancaster and E. G. Fayen, *Information Retrieval On-Line* (Los Angeles: Melville, 1973).
3. *ASIDIC Survey of Information Center Services.*
4. Organization for Economic Cooperation and Development, *Inventory of Major Information Systems and Services in Science and Technology* (Paris: OECD, 1971).
5. John H. Schneider, Marvin Gechman, and Stephen E. Furth, *Survey of Commercially Available Computer-Readable Bibliographic Data Bases* (Washington, D.C.: American Society for Information Science, 1973).
6. Vivian S. Sessions, ed., *Directory of Data Bases in the Social and Behavioral Sciences* (New York: Science Associates/International, 1974).
7. Lorraine Borman and Benjamin Mittman, "Interactive Search of Bibliographic Data Bases in an Academic Environment," *Journal of the American Society for Information Science* 23:164–71 (May/June 1972).
8. Harry B. Back, "What Information Dissemination Studies Imply Concerning the Design of On-Line Reference Retrieval Systems," *Journal of the American Society for Information Science* 23:156–63 (May/June 1972).