

# Providing Bibliographic Services From Machine-Readable Data Bases— The Library's Role

Richard DE GENNARO: Director of Libraries, University of Pennsylvania, Philadelphia.

*Libraries will play a key role in providing access to data bases, but not by subscribing to tape services and establishing local processing centers as is commonly assumed. High costs and the nature of the demand will make this approach unfeasible. It is more likely that the library's reference staff will develop the capability of serving as a broker between the local campus user and the various regional or specialized retail distribution centers which exist or will be established.*

This brief paper will attempt to counter the widely held view that the larger research libraries will soon need to begin subscribing to the growing number of data bases in machine-readable form and providing current awareness and other services from them for their local users.\* It will speculate on how this field might develop and will suggest a less expensive and more feasible strategy which libraries may use to gain access to these increasingly important bibliographic services. The key question of who will pay for these new services, the user or the institution, will also be discussed.

While it is clearly outside the scope of this paper to review the state-of-the-art of data base services, reference to a few key works and a brief introduction to the subject may be helpful.

The most comprehensive and authoritative review of the state-of-the-art of the field and its literature is the excellent chapter entitled "Machine-Readable Bibliographic Data Bases" by Marvin C. Gechman in the 1972 volume of the *Annual Review of Information Science and Technology*.<sup>1</sup> A useful selection of readings is *Key Papers on the Use of Computer-Based Bibliographic Services* edited by Stella Keenan and published jointly

\* This paper was developed from a talk by the author on a panel entitled "Library Management of Machine-Readable Data Bases." The Program was jointly sponsored by COLA, ISAD, and ACRL and took place at the ALA Conference in Las Vegas, June 24, 1973.

by the American Society for Information Science and the National Federation of Abstracting and Indexing Services in 1973.<sup>2</sup> *A Study of Six University-Based Information Systems* made by the National Bureau of Standards is essential and contains in convenient form comparative and descriptive information about these pioneering centers which are sponsored by the National Science Foundation.<sup>3</sup>

Some of the most useful and important data bases available are those that have been developed by the indexing and abstracting services as by-products of their efforts to automate the production of their regular printed publications. Like the publications, the tapes come in a wide variety of incompatible formats. Among the important producers are: Chemical Abstracts Service, BioSciences Information Service, Engineering Index Inc., American Institute of Physics, and the American Geological Institute. CCM Information Corporation (PANDEX) and the Institute for Scientific Information are two examples of major commercial suppliers.

Several of the scientific societies received substantial grants from the National Science Foundation and other sources in the 1960s for this automation effort, and it was generally expected that an important new market for the by-product tapes would develop among researchers in universities and in industry. Imaginative and forward-looking librarians and computer people at various universities applied for and received grants to establish centers where these new data tapes could be used to provide current awareness and retrospective search services to users. The National Aeronautics and Space Administration established a network of Regional Dissemination Centers at six universities, including the Universities of Connecticut, Indiana, and New Mexico, the North Carolina Science and Technology Research Center, University of Pittsburgh, and the University of Southern California. The National Science Foundation has been supporting centers at the University of Georgia, Lehigh University, University of California at Los Angeles, Ohio State University, and Stanford University. Other centers have been established at the Illinois Institute of Technology Research Institute and the University of Florida. It is worth noting that nearly all centers provide services free to their own institutional users and continue to be heavily subsidized. All seem eager to expand their markets to include paying customers from a larger region.

The latest entry into this field is the New England Board of Higher Education's Northeast Academic Science Information Center (NASIC) sponsored by NSF. NASIC's approach is basically different from the unitary centers that have been named. It will attempt to become a broker between the various existing centers and its own members, facilitating their access to existing services elsewhere. It will serve a ten-state region and is expected, perhaps somewhat optimistically, to become self-supporting after the three-year grant period ends.

The number of data bases available in the United States is now over a hundred and is growing rapidly, apparently without benefit of firm stan-

dards. A parallel development is taking place in Europe. As the number of available data bases increases, and as the activity at these centers expands, more and more librarians become interested in and concerned about how they are going to provide these new, important, and expensive services on their own campuses.

Interest among librarians in data base services is running high. A session at the Association of Research Libraries Conference in the spring of 1973 was devoted to it, and a program at the annual meeting of the American Library Association in Las Vegas on the subject was jointly sponsored by the COLA Discussion Group, the Information Science and Automation Division, and the Association of College and Research Libraries. While this interest is commendable and should be stimulated, it is also important that it be tempered and put into perspective by a realistic consideration of some of the costs and problems involved in providing these services. This is what the remainder of this paper will attempt to do.

The title of the ALA program was "Library Management of Machine-Readable Reference Data Bases." Implied in that title are two basic assumptions that are widely accepted: one is that libraries will play a key role in providing access to information in machine-readable data bases on their campuses. The other is that in order to provide this access they will have to acquire and maintain these data bases and develop the capability of searching and manipulating them for their local users.

The first assumption is valid; libraries will be responsible for assisting users in gaining access to information in this new form. The second assumption is highly questionable, if not invalid. It is extremely unlikely that many individual libraries will be able to afford to establish centers to acquire and process these machine-readable data bases. While it may appear that a straw man is being set up that can be easily demolished, the idea that academic libraries must and will begin acquiring and servicing many large and expensive data bases, and even statistical data banks, is still widely enough held that it ought to be put to rest.

How did this idea gain such currency? Perhaps it was because the first available data bases were from the indexing and abstracting services and contained machine-readable versions of their printed indexes. Since libraries subscribed to the printed editions, it followed that they should also subscribe to the tape editions. The same is true for the census tapes. Libraries were the chief repositories for printed census publications, so it was natural to assume that they would have to subscribe to and make available the machine-readable census data as well. We now know better about the census tapes; the problem was simply beyond our resources, and they are being made available from specialized centers. A similar solution may well emerge for the bibliographical data tapes of the indexing and abstracting services.

To help put matters into perspective, it might be useful to review a few other ideas we had in the last two decades on how certain technological de-

velopments would be implemented in the library. Take microfilm, for example. Back in the 1950s when microfilm came of age for library use, many librarians thought that every major library would require its own laboratory where large quantities of film could be produced and processed under the direction of a new breed of librarian called a documentalist. Several major libraries did establish such laboratories for a time, but the only remaining ones of any significance are at the Library of Congress and a few other large libraries. Most of the others were put out of business by the copying machine, the local service bureau, and commercial micropublishers—and the documentalists became information scientists.

Library automation provides other interesting examples. Many of us recall that in the 1960s it was a commonly held view that each major library would have to automate its operations, and that librarians would learn to master the computer that was soon to be installed in every library basement, or see themselves replaced by computer experts. As we all know, it did not happen that way. Librarians will probably end up with computer terminals or minicomputers, with software packages supplied by library cooperatives or commercial vendors.

When the MARC tapes were first made available, it was assumed (and this is what the MARC I experiment was all about) that each library would have to subscribe to the tapes and design, implement, and operate its own system to use the data in its cataloging operations. Again, it did not happen that way. MARC data are being used by libraries, but indirectly through cooperative centers such as OCLC, or through commercial vendors of card services such as Information Design or Josten's, Inc. Individual libraries are not subscribing to MARC tapes, as we had thought would be the case.

The point of citing these few examples is to suggest that it is extremely difficult in the early stage of a new technology to predict with any confidence how it will be introduced and implemented, and what effects it will have. We seem to have a natural tendency first to try to cope with each new technological development on a do-it-yourself individual library level, and when experience teaches us that implementing the particular technology is more difficult and more expensive than we thought, we regroup and try a broader-based approach. This is approximately where we are with data base services; it is time for a broader-based approach.

Again, it is unlikely that libraries will provide access to machine-readable data by setting up their own campus information centers to acquire and process data bases. Anyone who takes the time to look at a list of data bases available and their annual subscription rates will understand that research library book budgets will not be large enough to cover these additional subscription costs. In fact, the subscriptions are only a minor element in the total cost of providing these services. The data bases must be cumulated and maintained. Programs to manipulate and access them in their many nonstandard formats and contents must be written or adapted.

The cost of administering and marketing the services and interfacing with the users will be high.

Perhaps the most critical question to be answered is: will the individual user be charged for the services he uses or will the costs be absorbed by the university? The answer to that question will determine how and to what extent the machine-based services will be used in the future. If they are offered free, as are traditional library services, then one can assume with some confidence that a substantial demand for them will materialize. This has in fact been the early experience of the centers at the University of Georgia and Ohio State and others where use has been totally subsidized by grant money.<sup>3, 4</sup>

On the other hand, if the individual user is asked to pay for these services out of his own pocket or even out of departmental or grant funds, the market for them will be severely limited. It is extremely unlikely that large numbers of faculty and other researchers in universities will be seriously interested in becoming *paying* users of machine-based information services. The experience of C. C. Parker at the University of Southampton may prove to be typical.<sup>5</sup> He reported a drop from forty-seven to five users of an SDI service after charges were introduced. It was not that the users could not pay the charges, but that they preferred to use their resources for other more important needs. The National Library of Medicine recently instituted user charges in the MEDLINE system in order to effect a needed reduction in the number of users.

The case for giving these services to users free is theoretically sound in the traditional library context, but there are practical difficulties. First, these services will be expensive and they will require a net addition to library budgets rather than a transfer from one activity to another; the prospects for such budget increases seem dim in the next few years. Second, if the services are offered free, there will be no natural or automatic mechanism for controlling their use, and such control is essential to limit costs. Once users get on a free subscription list they will tend to stay on it whether they actually use the products or not. This happens in many libraries where current accessions lists are regularly sent to faculty, most of whom discard them unread. On the other hand, there is ample precedent for charging a modest fee for certain services in libraries. The best example is the almost universal charge for photocopies. In those instances where libraries offered free copies, the service was abused and charges had to be reinstated.

It seems likely that a combination of institutional subsidy and individual charges will evolve as the dominant method of paying for machine-readable services. In order to recover some costs and prevent abuses, an appropriate system of charges will have to be instituted in spite of the logic of the argument for free services. Incidentally, the case for free computer time in universities is perhaps equally valid, but it has never been accepted by the responsible budget officers.

Regardless of who pays, these services will have to be advertised and marketed aggressively to reach the limited number of potential users on each campus. It will not be enough to announce their availability and wait for customers. But even the best salesman on the most research-oriented campus will probably fail to find enough users to justify the high costs of providing the extensive and diverse subject coverage that every university will require. The solution, of course, lies in the establishment of a small number of comprehensive regional or even national information processing centers, possibly backed up by a much larger number of specialized centers or services for particular subject or mission-oriented fields such as physics, chemistry, medicine, pollution, urban studies, census data, etc.

Libraries will play a key role in facilitating access to data bases by functioning as the interface or broker between the users on campus and these regional and special processing and distribution centers. This means that they must develop a new kind of information or data services librarian on their reference staffs whose function it will be to publicize these services and maintain extensive files of information on their scope, contents, cost, and availability. These reference specialists will also guide users to the most appropriate services, help them to build and maintain their interest profiles, and provide assistance with the business aspects of dealing with vendors.\*\* After an initial start-up period, this function should and doubtless will become a fully integrated part of the regular reference service, and the need for specialists will disappear as this knowledge becomes a part of every reference librarian's repertoire. The available data base services fall into two main categories: off-line batch and on-line interactive services. The most commonly available up to now have been regular off-line current awareness (SDI) services based on an interest profile; these have been supplemented by occasional requests for retrospective searches of the older files. The results of these off-line searches are delivered to the subscriber by conventional mail. On-line services permit the user or the reference specialist to access a portion of the data base directly via terminals and telephone lines and perform the search in an interactive mode. Some results are immediately displayed on the terminal and others are sent by mail.

The Lockheed Information Retrieval Service and Systems Development Corporation have recently begun offering interactive searching with on-line computer terminals of a large selection of the most useful bibliographic data bases. With this capability commercially available from leased terminals on a fee-per-use basis, it will be difficult for a university or even some existing centers to justify subscribing to and maintaining these data bases for their own limited use. If Lockheed, SDC, and other vendors can develop the market and operate these services at a profit, they may be able to

\*\* The University of Pennsylvania library recently established a Data Services Office based on this concept with encouraging early results.

satisfy a very substantial portion of the need for these new bibliographic services.

MEDLINE, TOXLINE, RECON, and the New York Times Information Bank provide other models for specialized and centralized interactive services. Some authorities assert that this trend toward on-line interactive searching will accelerate and eventually supersede tape searching.<sup>6</sup> Others argue that the cost of maintaining and searching on-line the really large data bases is prohibitive and will remain so for several years to come. It seems most likely to this author that the trend will be toward on-line systems covering a limited period of time, probably the latest three to five years, with supporting off-line services for retrospective searches. If this proves to be the case, libraries will find it practical and convenient to make terminals available at or near reference desks.

A close look at the several centers which now exist on individual campuses would probably show that they are heavily subsidized by grant or other outside funds, and that they are trying to expand to serve their states or even wider regions in order to achieve greater cost effectiveness. These centers deserve the credit that is always due pioneers. They are in the process of developing the patterns for providing these services in the future. One of the chief lessons they may have already taught us is that a single university, or even possibly a single state or region, is not a large enough market base upon which to build this activity. These centers will require a large volume of business to justify their high overhead and operating costs and they will seek and welcome additional paying customers.

To summarize and conclude, libraries will play a key role in providing access to machine-readable data bases, but they will generally not do it by acquiring and managing these data bases in local campus centers because of the high costs involved. These high costs and the limited market will restrict the number of processing centers to several regional or even national centers, supplemented by a larger number of specialized discipline and mission-oriented services. Many data bases and services will be available on a fee-for-service basis either through existing centers or directly from professional societies, government agencies, and commercial vendors with the library serving as facilitator or broker. It seems likely that a combination of institutional subsidies and individual charges will emerge as the pattern for paying for these new computer-based bibliographical services.

#### REFERENCES

1. Marvin C. Gechman, "Machine-Readable Bibliographic Data Bases," in *Annual Review of Information Science and Technology*, v. 7 (Washington, D.C.: ASIS, 1972), p.323-78.
2. Stella Keenan, ed., *Key Papers on the Use of Computer-Based Bibliographic Services* (Washington, D.C.: ASIS, 1973).

3. B. Marron, and others, *A Study of Six University-Based Information Systems* (Washington, D.C.: National Bureau of Standards, 1973 [NBS Technical Note 781]).
4. James L. Carmon, "A Campus-Based Information Center," *Special Libraries* 64:65-69 (Feb. 1973).
5. C. C. Parker, "The Use of External Current Awareness Services at Southampton University," *ASLIB Proceedings* 25:4-17 (Jan. 1973).
6. M. Cerville, L. D. Higgins, and Francis J. Smith, "Interactive Reference Retrieval in Large Files," *Information Storage and Retrieval* 7:205-10 (Dec. 1971).