GLYN T. EVANS
Director for Library Services
State University of New York
Albany, New York

Bibliographic Data Centers for New York State

In the years which have elapsed since librarians began to use the computer to attack some of the problems they face, massive technological change has altered computer design and performance. Technical development has resulted in larger, faster machines, with storage capacities and processing speed factors higher than a decade ago, and a consequent dramatic reduction in the unit cost of storing and processing data. Simultaneously, teleprocessing has also developed rapidly, with programmable CRT terminals relatively commonplace, where little more than a decade ago, in 1962, a computer manufacturer announced the first linkage between a Telex and a computer.

As a result, the concept of a large computer file of bibliographic data constantly maintained and usable at random by a number of users simultaneously (or rather concurrently) for a variety of purposes (much like a card catalog in a large library) has become not only technically feasible but fiscally desirable. In this paper I want to describe the work being undertaken in New York State to realize this concept.

Environment

I will provide some sense of the size of the problem. Firstly, the state of New York is rich in library resources. One figure commonly quoted—and certainly inaccurate but nonetheless an indicator—is that 10 percent of the bibliothecal wealth of the country is located in New York State. There are some 217 academic libraries in the state, including the great collections at Cornell, Columbia and New York University, for example. Seventy-two of the libraries are on campuses of the State University of New York; there are

nineteen campuses of the City University of New York, some jointly adminis-
tered with state university; and the remainder are private institutions. There
are extensive public library systems, including such major resources as New
York Public Library, and Buffalo and Erie County Public Library. The State
Library in Albany is a fine library; and there are good school library services.
Finally, there are magnificent special and private libraries, including such
organizations as the Research Libraries of New York Public Library, the
Engineering Societies Library, and libraries of the large industrial concerns.

I hope I will be forgiven if I speak of the libraries of the State
University as a rich and diverse collection in themselves. They include the four
university centers at Albany, Binghamton, Buffalo and Stony Brook; fourteen
four-year colleges of arts and science; four medical schools; six specialized
colleges including the agriculture, veterinary and industrial labor relations
schools at Cornell; forestry and environmental science at Syracuse; ceramics,
maritime and optometry colleges; and six agricultural and technical colleges.
Finally, there are thirty-five community colleges administered jointly by the
State University and the local communities.

Secondly, there is already within New York State a strong tradition of
library cooperation. The public libraries are grouped into twenty-one cooper-
ating systems. The State Education Department Division of Library Develop-
ment designed and supports nine regional reference and research councils
(3Rs councils), including METRO in New York City, which completely cover
the state and are a local mechanism which bring together all types of libraries
for their mutual benefit. State University is itself grouped into four regions
and the librarians in those regions are beginning to explore local modes for
cooperation. Finally, there are other groupings, such as the Five Associated
University Libraries (FAUL) which comprises Binghamton, Buffalo, Cornell,
Syracuse and the University of Rochester, and has now added as affiliate
members Albany and the State Library, and the Associated Colleges of the St.
Lawrence Valley.

Thirdly, the state has demonstrated considerable support for libraries,
such support manifesting itself in the NYSILL interlibrary loan system, the
work on the ANHLTS processing center project, and aid for public libraries.

Bibliographic Data Centers

Having described, albeit too briefly, the libraries in New York State, I
should now concentrate on the latter half of my title and define bibliographic
data centers. A bibliographic data center is a central computer system which
provides bibliographic data, primarily to libraries but also to their users, and
whose prime operational mode is by on-line interaction supported by batch

services. The data stored and supplied will range from item location data through standard bibliographic data to index and abstract data. A bibliographic data center will *not* involve itself with handling and physically processing library materials.

I prefer to visualize and discuss the role of such a center from the point of view of the librarian who will use the services offered by it.

The librarian needs, close at hand, a range or first circle of files (see figure 1) to control circulation, acquisitions and in-process items, serials check-in, and probably union location lists of serials and monographs. These specifically job-oriented files are the volatile areas of library record keeping where there is the greatest change of state in the files, and where currency of information is vital.

While these files need authoritative, standard data, they do not need a full standard bibliographic description. Instead, each record will be composed of a selection of data elements, the selection being based on the absolute necessity of the presence of each element in order to be able to perform each task. For example, a subject heading is not necessary in a circulation record which would be accessed by name, title, call number, etc.

Just as the files are shown to be job-oriented, so will it be clear that the librarians must be able to work with each file and change it (and the records in it) by adding, changing or deleting data. This statement, obvious enough, raises interesting questions of file security.

Each file is shown to be linked to all others since data will clearly move between some of the files at this level, although perhaps not among all the files.

In the second circle of files (figure 2), the librarian requires access to authoritative bibliographic data primarily in order to describe the collection in hand, but also for verification and searching purposes. The data will clearly be MARC-formated data for monographs, serials, maps, and so on. In addition to relying on the Library of Congress to supply data, each library will accept the responsibility of contributing data to the common data base in a shared cataloging mode, as do members of the Ohio College Library Center.

The records in this set are not as volatile as in the first group. In fact, once a record is established on the file there should be no change other than that which emerges from normal upgrade through usage of the record, and the free proofreading and scrutiny which such usage engenders. It is true, however, that records may be supplanted in toto by other versions, particularly if a system of preferred cataloging source (or sources) is established.

This does not mean that standard bibliographic data will be forced down users throats, but merely that a user will adopt and adapt them for local usage while the master file is protected. While the file will be used heavily, it will
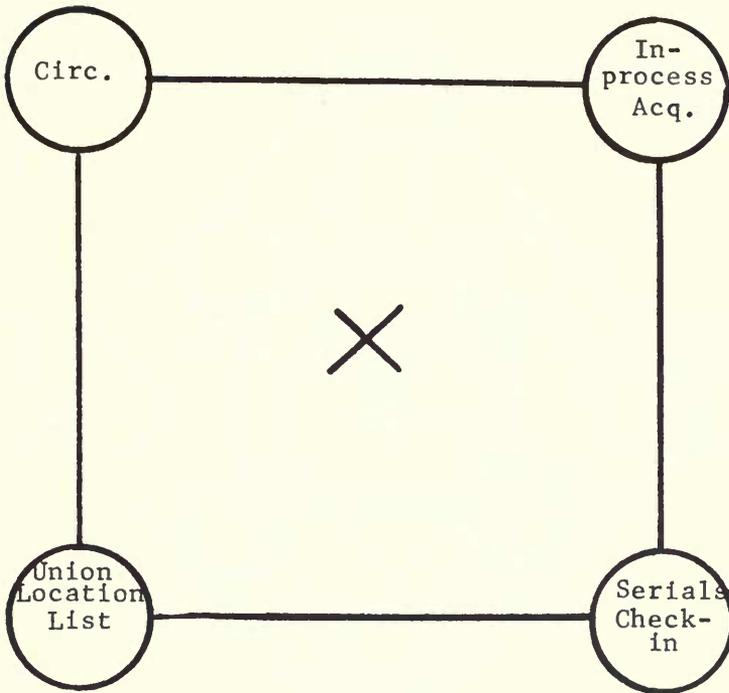
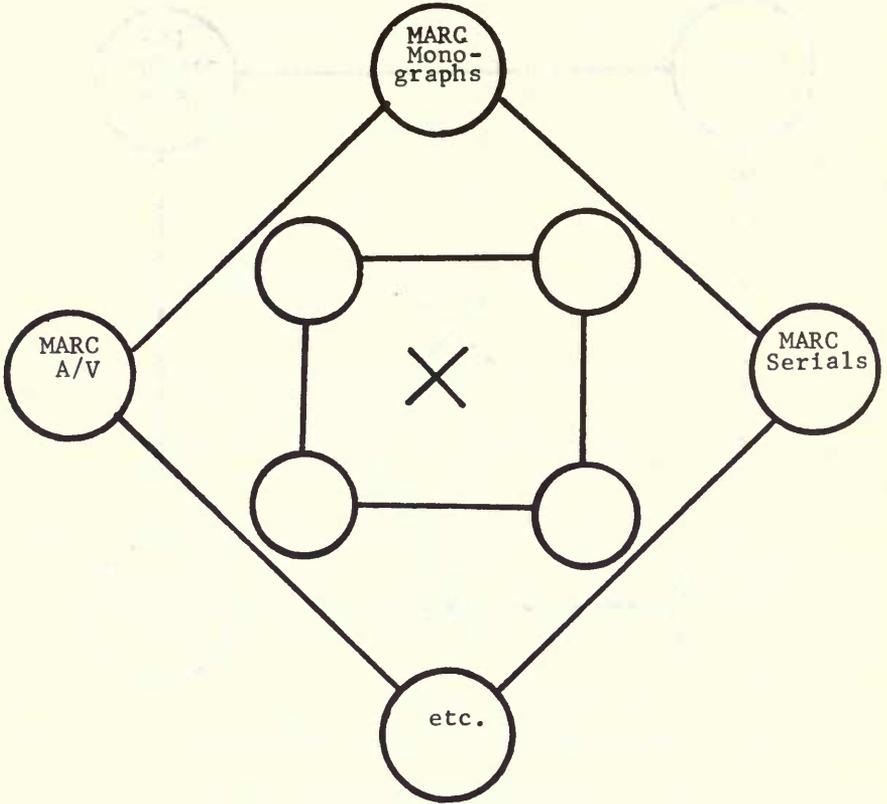Fig. 1. The First Circle of Files: Volatile Working Files

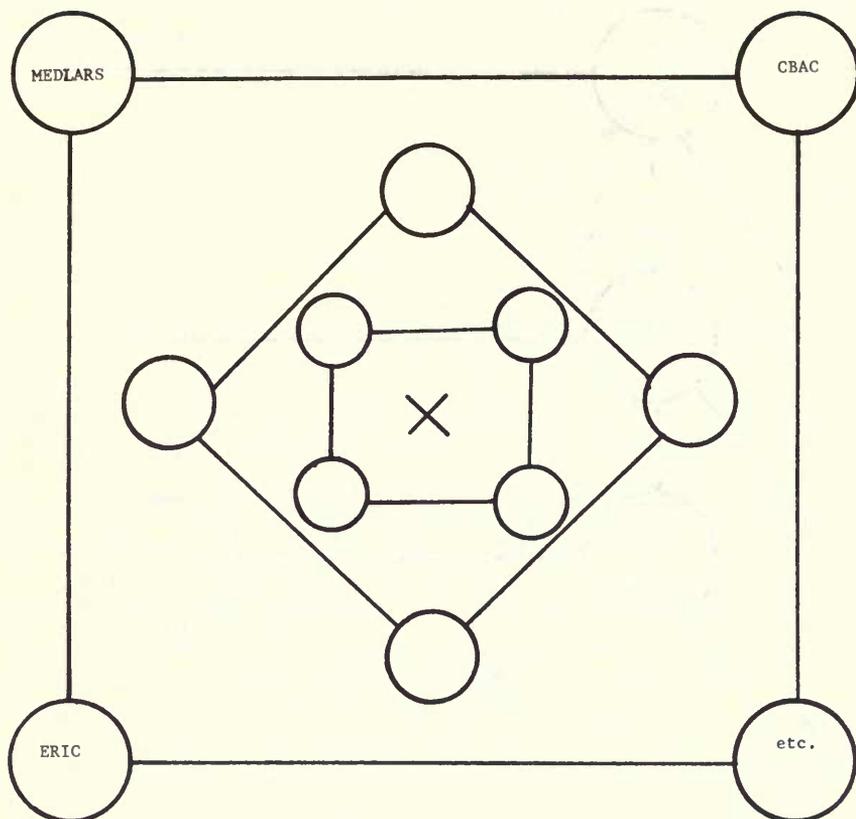Fig. 2. The Second Circle of Files: Standard Bibliographic Data

Fig. 3. The Third Circle of Files: Abstracts and Indexes

not be volatile (in the sense that there is no activity against the records, only retrieval), but it will grow rapidly as more and more libraries contribute to the master file.

From the third and outer circle of files (figure 3), the librarian, on behalf of the user, and indeed the user directly, will acquire the ability to search data bases comprised of abstracting and index data, such as MEDLARS, CBAC, BA Previews, etc. These files will be used for retrospective searching in both interactive and batch mode, and will also be used for selective dissemination of information services.

The files will be very large, but at the same time the most stable, inasmuch as the data themselves will not change (indeed some may be
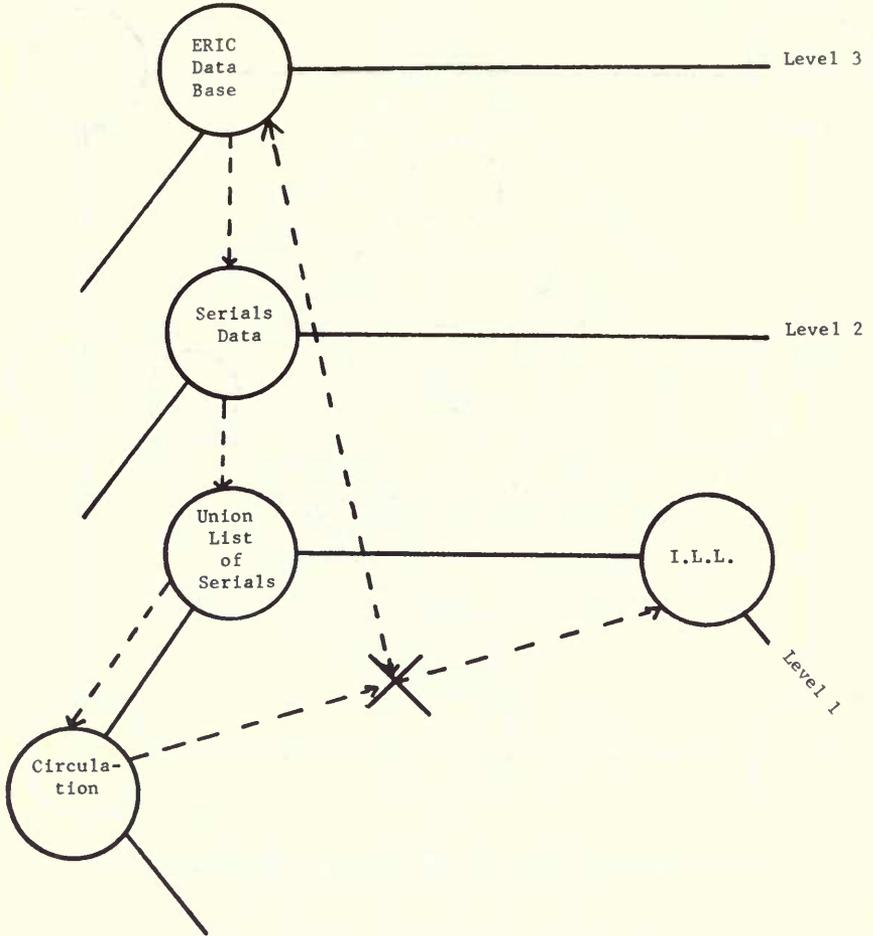
Fig. 4. Linkage Among Levels: A Subject Search

protected from change by copyright agreements). The size of the on-line files, and the proportion available for retrospective search, will be a reflection of the volatility of the subject matter; files of scientific information will probably cover a shorter time span than social science or humanities files. The reader should note that although in the three figures only four files are shown at each level, this number is only used to illustrate the concept. In the

seventh *Annual Review of Information Science and Technology*, the review chapter on machine-readable bibliographic data bases lists many data bases.[1]

While the existence and availability of these files is essential, even more important is the creation of the necessary linkages among the levels and among files at the samel level. Two examples will suffice to indicate what the linkages would do in an on-line system and why they are essential.

In figure 4 the librarian (or user) is searching a data base, say, ERIC, and retrieving a list of citations. These citations are really almost useless unless they carry the user to the document. (Giving the user citations has not solved his problem; it has merely changed it.) In the figure the serial citations are passed through the serials data file to impose bibliographic consistency, although this may not be a necessary step in all cases. From there, the citations are passed to a Union List of Serials data base to discern locations. From there they go to a circulation file to assess availability and then back to the librarian or user, some data perhaps being forwarded into an interlibrary loan module either directly or following user decision. I suggest, of course, that this be automatic and that the results of all searches be passed through this route, such that the user gets a complete report in response to his search request.

Figure 5 is an example of a different kind of linkage, one devoted primarily to internal housekeeping, which shows the path of events following the decision to order a new serial. The librarian goes out to the file of standard bibliographic data to acquire authoritative data prior to placing the item in the on-order acquisition in-process file. The subsequent receipt of the first of the serials would trigger a range of actions. The in-process file would be amended; full serials control of data would be needed to update the serials catalog; an addition would be made to the Union List of Serials; and a serials check-in record created, again based on data already extant in the serials data base and in-process files; and finally data from the serials check-in record would be needed both for the serials catalog and the circulation system.

Figure 5 differs from figure 4 in that there is much delay between some steps rather than immediate transfer of information in order to complete the transaction.

Thus far these files have been depicted in a flat time-dependent schema in the relationships of one to the other, but there is another set of relationships which are equally important—their relationship in space.

I do not believe we yet know which is the most appropriate level at which all of these files should be maintained. Should circulation be done at each local campus, by subregions within state, or statewide? At the other end
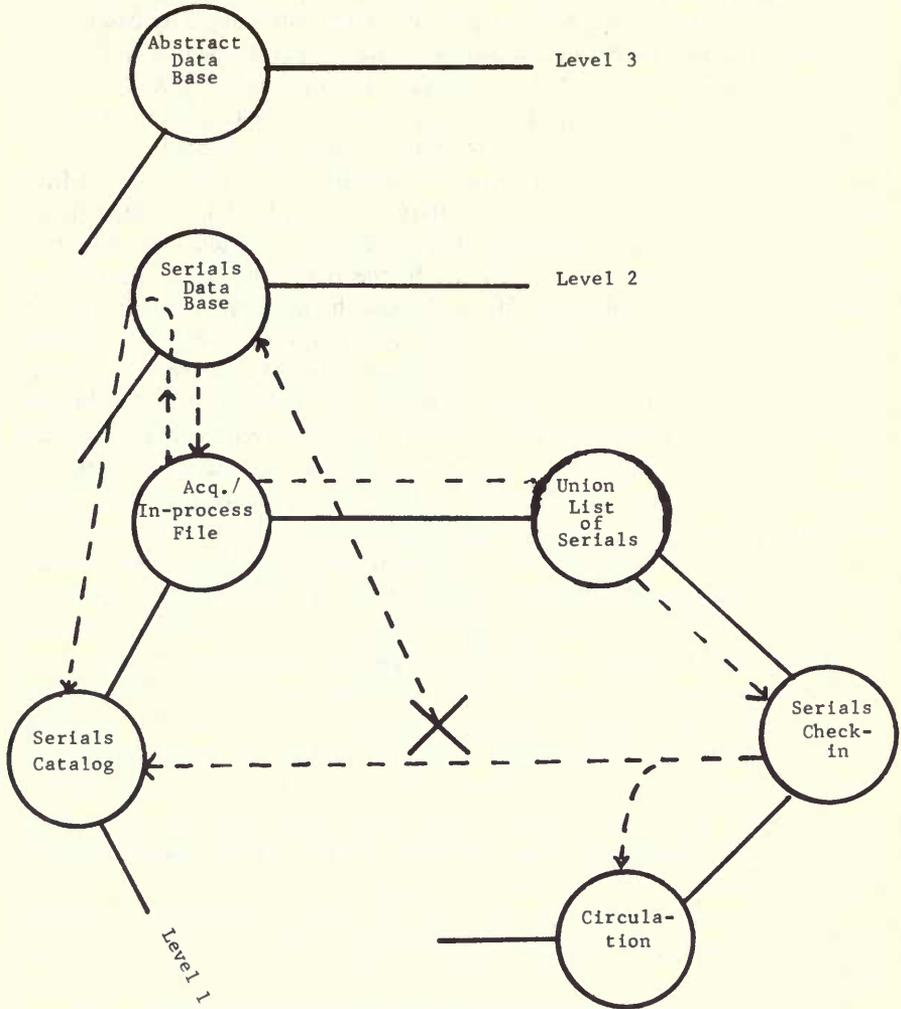
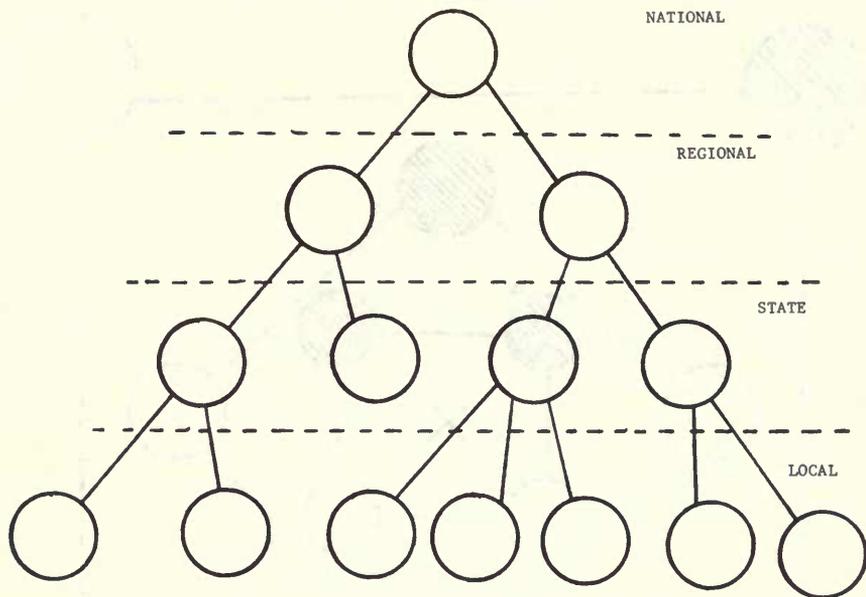Fig. 5. Linkage Among Levels: Order a New Serial

Fig. 6. Levels of Appropriateness

of the scale, how many centers should be carrying, for example, engineering or geographical data—one national center, a set of regional centers, or every campus?

The answers to these questions will lie in the usage to which each data base is put (measured quantitatively) and whether or not linkages can be forged between the present disparate systems which exist on a national basis. There must be a level at which each system can achieve close to optimal cost benefit for its users.

When figures 3 and 6 are imposed one on the other, a three-dimensional picture, more like a contour or relief map, emerges where files will be located at varying levels and distances from the user, and from each other.

The important, indeed vital, fact to remember is that *where* the file is held is of no concern to the librarian. What is important is *access* to the file. It is possible to set up performance criteria that insure good response times (measured in seconds) for the functions described above. That is what large on-line interactive bibliographic computer systems are all about. Furthermore, let me reiterate that these systems will be cost-beneficial to the participating library.
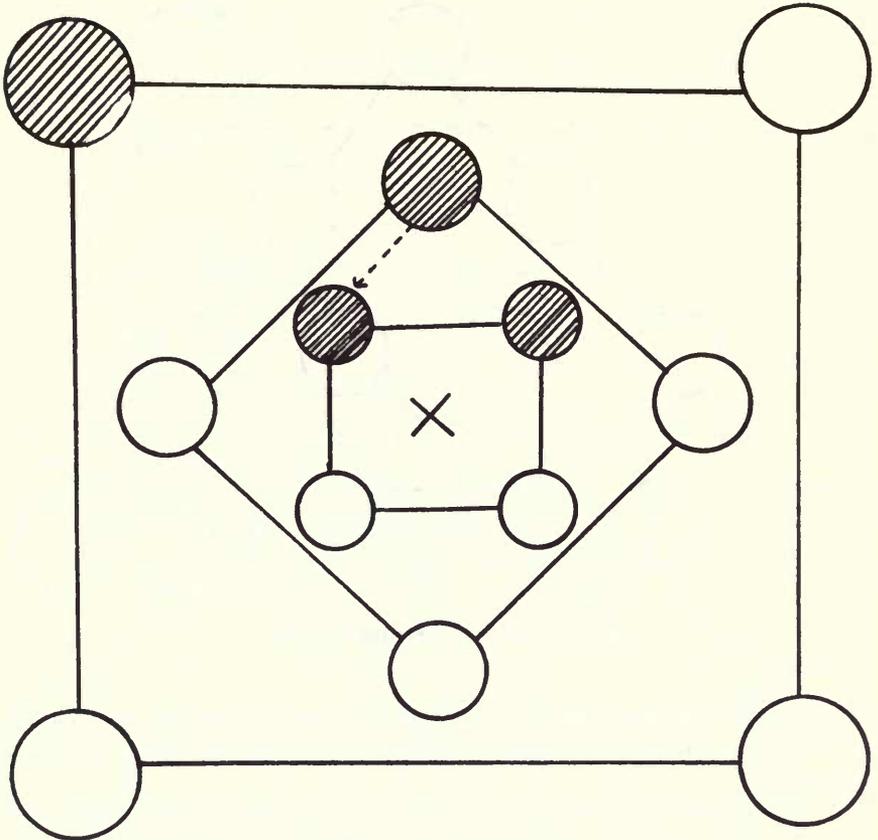
Fig. 7. Modules Which Already Exist

Work in New York State

With this bibliographic data center in mind, what progress is being made toward the realization of such a system in New York State? First let me admit that the description of the center has that all-too-familiar, all-too-depressing ring of blue sky about it. We have been reading and dreaming about similar systems for a decade.

I suggest, however, that we are a lot closer to realizing such a center than is generally appreciated. In figure 7 I have filled in examples from each level to indicate that systems do exist for each.

In the first circle one can point, for example, to the Syracuse University on-line in-process and shelflist system and its circulation module. The Ohio State University Libraries Control System, described by Hugh Atkinson,[2] is another example.

The Ohio College Library Center is performing the function of the second circle as it provides on-line access to shared cataloging data, presently for monographs, but hopefully for all other media. It should also be remembered that a link does already exist between these first two levels when OCLC provides OSU with a weekly cataloging tape to maintain their own (OSU) on-line shelflist for the circulation system.

Examples of the third level are the State University of New York's Biomedical Communication Network and the National Library of Medicine's MEDLINE system, through both of which the MEDLARS data base may be interrogated interactively by librarians or users.

Thus in New York (on the SUNY central computer) we already have a segment of the outer level. And clearly if the system can interrogate MEDLARS it can also work against other data bases. We are, therefore, actively examining other data bases and working on a schedule with which to bring up other data bases, ERIC and BA Previews, for example.

At the inner level, State University has taken the OSU libraries Control System and is adapting it for use on the SUNY central computer to serve in the first instance the Albany University Center Campus, by September 1973, but we plan to amend the programs to handle multiple files and gradually extend the use of the system throughout the university. Several campuses of the university have already expressed keen interest in participating in such an expansion. This is a very attractive mode of action for another entirely different reason, namely that the result of such action would be a start on the development of a University-wide shelflist, a tool which is absolutely essential for the best utilization of the library resources of the university. In addition, also at the inner level, the NYSILL interlibrary loan system runs partially in an interactive mode.

What of the second level? Work in this direction is taking very positive form within the state. State University and the State Education Department have within the last few months created a Task Force on Library Data Centers, which has eighteen members. Six of these members are from State University; the State Library and the Division of Library Development are each represented, as are three of the 3Rs Regions. Other members are from Five Associated University Libraries, New York University, New York Public Library, Columbia University, City University of New York, ANYLTS, and the Commission of Independent Colleges and Universities of the State University of New York.

This group has met approximately each month since its first meeting on October 17, 1972, and has prepared the following set of guidelines, objectives, and recommendations.

OBJECTIVES

The mission of the Task Force on Library Data Centers is to promote the development and to advise in the implementation and operation of a statewide bibliographic computer network which will improve services to users of libraries within the state through on-line access to various data bases for the purposes of shared cataloging, circulation control and interlibrary lending, serials control, union lists, information retrieval, acquisitions and other similar services.

GUIDELINES

1. The Task Force will be guided by such technological advances and developments as have already been made, or are being made, in the fields of library data processing, recognizing that the use of existing technology will generally be more cost beneficial than parallel or competitive development.

2. The above statement, however, shall not inhibit independent development where it is considered that such development will be clearly cost beneficial to the libraries of the state.

3. The Task Force will seek to make available to libraries within the state such data bases as are available where access to such a data base will clearly be of cost benefit to users. These data bases may be of cataloging data, holdings data, indexes and abstracts.

4. The Task Force will encourage the acceptance of and adherence to appropriate national standards for data input, format, and description. Such systems as are developed will interface with other bibliographic data networks, such that a national network can be realized. Developments made by the system should be made available to other networks on a reciprocal basis for the purpose of contributing to the total network by interlocking with other developments.

RECOMMENDATIONS

Within the above guidelines, the Task Force makes the following recommendations:

1. That capabilities and services such as those provided by the Ohio College Library Center be made available in New York State, as a first step toward

providing the above bibliographic services. It is recognized that the needs of libraries may require more than one center.

2. Planning and development of such capability and service should be initiated immediately, and should proceed as far as possible pending the availability of funding. Concurrently, a continuing evaluation of best methods for total fulfilment of the Task Force objectives should commence.

3. That the State Education Department and the State University jointly seek state and other funding for planning, development and implementation, but that operating costs for the provision of services are expected to be supported by participating institutions.

4. That such services will be available to private and public academic libraries, and to such other users as feasible.

5. That the planning and execution of all studies and development work will be the joint responsibility of the State Education Department and the State University with the advice of the Task Force.

6. When operational, the bibliographic data center will have governance that provides for appropriate participation by public and private institutions, and SUNY and the State Education Department.

****

The Task Force is now seriously examining the opportunity for developing a system at the second level, either by totally replicating the SIGMA V System used at OCLC or, much more likely, performing a functional replication of the OCLC system on hardware which exists within the systems, and using existing (and expensively trained and experienced) personnel. In particular, State University, State Education Department and New York Public Library have available enough experienced people to build a very powerful team. Whatever path is chosen, there is clear intent to develop a system and, more important, clear fiscal support for such work.

It would have been pleasant and very gratifying to report that the work is done, rather than to present an account of work in progress. However, we are sufficiently along the road on which our thinking is taking us as to allow hope that some worth can be gleaned from this report.

New York cannot do the work on its own, of course. We are already drawing heavily on work performed elsewhere, and will continue to do so. We do hope that our work is a contribution to a national on-line bibliographic

data network such as must be realized in the near future; indeed we are committed to interface with and share in such a development, even though at present we may barely be able to enunciate the problems which will be encountered in such a network development.

If we believe anything concerning our work in New York, however, it is that the job can, should, and will be done.

REFERENCES

1. Gechman, Marvin C. "Machine-Readable Bibliographic Data Bases." *In* American Society for Information Science. *Annual Review of Information Science and Technology*. Vol. 1. Chicago, Encyclopaedia Brittanica, 1972, pp. 323-78.

2. Atkinson, Hugh C. "The Ohio State On-Line Circulation System." *In* F. Wilfrid Lancaster, ed. *Proceedings of the 1972 Clinic on Library Applications of Data Processing: Applications of On-Line Computers to Library Problems.* Urbana, Ill., University of Illinois Graduate School of Library Science, 1972.