
Options for Cooperative Support of Access to Numeric Files

by *Jean Slemmons Stratford*¹
Institute of Governmental Affairs
University of California, Davis

Introduction

Jim Jacob's work on levels of service and levels of access provides an excellent starting point for exploring options for cooperative support of access to numeric files². Jim divides types and levels of service into four basic categories. In the first, General Data Services, he delineates the full range of services that an institution might offer in connection with machine-readable data. Services are laid out in a hierarchical manner. The three remaining lists, Library Data Services, Reference Data Services, and Computing Services, outline services that fall into each of these more specialized categories. While he does not directly address cooperative support, he outlines quite clearly the possible levels of support and service. He also makes the equally important point that it is not necessary — and probably not desirable — for a library to attempt to provide “full service” for machine-readable information on its own.

That being the case, building partnerships to provide enhanced levels of service makes a great deal of sense for many libraries. However before setting out to forge these partnerships, a library must take stock and determine exactly what levels of support it can provide in-house, where to draw the line, and what sort of partnerships it might logically seek.

In the data archives community, there is no one model for the provision of data services—no right way to do it. Each data library or archive seems to have its own unique structure, procedures, and services. Therefore, traditional libraries entering into the data services arena will do well to review their goals and objectives and formulate a mission statement for data services. Such a statement requires deliberate managerial decisions on what levels of staffing and funding are available to commit to data services. It must also determine what levels of service are desirable given its mission and supportable given the resources at hand. From there, formal collection development and public service policy statements are appropriate and useful tools for communicating these decisions to the library's clientele. Once the library delineates the role it can and will play, it can seek partnerships that will strengthen and complement its services. Doing the latter without the former may result in difficulties when the objectives of cooperation are

unclear and the division of responsibilities and authority between cooperating organizations is ambiguous.

Just as there is no one right way to deliver data services, the right way for any given library will be governed to a great extent by its larger institutional context. A library department seeking to define its role in providing data services must understand its place within the library and the role of other actual or potential data service providers within that library. If the library services (or is considering servicing) datasets in other reference units, it should consider the pros and cons of establishing additional decentralized data services. There may be economies of scale in consolidating services where subject and/or technical expertise is strongest. There may also be an established philosophy within the library that dictates one approach over another.

The library must also understand and take into account its role in relationship to other organizations within the institution, as well as externally. For example, if a campus has other strong units with a history of established data services, the library will need to be aware of those units and their services when deciding its role. If the campus administration has funded other units to provide some types of data services (for example, GIS systems), the library may want to develop arrangements for housing and/or servicing any geographic files it acquires in those units rather than duplicating services available elsewhere.

Once it has completed its deliberations and come to some decisions on the levels of service it will provide, a library may want to review its selection of machine readable items. For example, the library may decide not to support files without their own extraction software. In that light, (without agreements to support them elsewhere on campus), the library would want to ensure that it had not selected items like the Current Population Survey or the American Housing Survey through the depository library program. Conversely, if a library is not selecting or acquiring files that another unit is willing to support, perhaps they should be acquired.

Collaborative alliances may take one or more forms. They may be purely informational and informal: they

may be for the sharing of expertise, information, or solutions to problems; they may also be more formal and involve a division of labor or resources in support specific files or classes for files. Anything other than the most informal of collaborations will benefit from a written agreement. Such a document can clarify many aspects of the arrangement. It should include information on what the aims of cooperation are, how the collaboration will work, what the division of labor will be, what each party's level of commitment- is in terms of resources, services, whether commitments are ongoing or for a set period of time, etc.

With these points in mind, here is an overview of some of the many possible sources for strategic alliances to enhance support of data collections.

The Campus ICPSR Official Representative

ICPSR, the Inter-university Consortium for Political and Social Research, is a consortium of nearly 400 institutions worldwide. One of the primary functions of the consortium is to support a central repository and dissemination service for machinereadable social science data. For many members, the primary benefit of membership in ICPSR is access to the consortium's vast data collections. Many of the data series held and distributed by ICPSR will be familiar to librarians in their printed forms. The ICPSR membership and data distribution is handled on each member campus by an "Official Representative." Currently, there is a trend toward housing the ICPSR membership and data collection within a support unit such as the member institution's library or computer center. However, historically ICPSR ORs have come from other areas as well. ORs include among their ranks not only librarians and programmer/analysts but teaching faculty in a variety of social science disciplines and academic staff from research institutes and programs. There is a substantial body of expertise in the organization in use of social science machine-readable data. In an institution where the ICPSR membership is handled outside the library, this would be an excellent first place to look for strategic alliances. However, the range of options in servicing library datafiles may be limited. Two options come most readily to mind: 1) informal collaboration and sharing of expertise, 2) expanding access to data sources by including the ICPSR collection in the Library's OPAC regardless of physical ownership and location of the collection. The latter has been done successfully at several institutions and a variety of approaches have been used.

Computing Facilities

As Jim notes, users of machine-readable information must have access to computing services. Jim divides these services into four basic categories: data storage services,

copying and subsetting services, data retrieval services, and data analysis services. Provision of even the most basic data storage services will require some access to appropriate hardware and software. Given the dramatically short life of computer products, computing services is one area where cost may quickly outstrip a library's resources. Therefore, the library will benefit from a clear understanding of what levels of service it can support in-house and what other institutional resources are available to provide computing services.

A library may acquire datafiles on a number of storage media from floppy diskette and CD-ROM to various and sundry tape formats. The range and type of media acquired will determine whether the computing facilities needed to support even basic data storage services are minimal or more extensive. If a library limits its acquisitions to diskettes and CD-ROMS, the equipment required to verify and backup datasets will be manageable. However, equipment and software must still be available and kept up-to date to perform these simple procedures.

For the other levels of services (copying, subsetting, retrieval and analysis), the equipment requirements escalate rapidly. Additional hardware and a broader range of software are required for these latter services. As the hardware and software requirements increase, the human resources that must be devoted to servicing the files are also dramatically increased. Clearly, this is an area where collaboration may be in order.

On a typical campus, there are several places to look for partnerships. Centralized computing facilities are a likely possibility and may have resources to commit to supporting access to the library's datafiles. Most such facilities are better placed than any library can hope to be for the simple reason that they have budgetary resources committed to maintaining and upgrading a volume of hardware and software. More formal collaboration with centralized facilities might include something as simple as providing end users with access to lab equipment and ensuring that the lab provides support for appropriate software packages for use with the library's datafiles. Greater collaboration might encompass shared access to and support of equipment, delegated support for data storage services (for example, the library acquires and catalogs the datafiles which are housed and retrieved in the lab), or the provision of copying and subsetting services to end users by referral.

Other computer labs may also be maintained by computing intensive departments, colleges or institutes. For example, the research emphases in many geography departments may make it feasible for them to maintain their own GIS labs. C)n my own campus, there is a

college-supported computing facility for social scientists. In these specialized labs, direct access to equipment by outside users may be more problematic. However, libraries will still benefit from informal ties with their personnel, as these staffs frequently have expertise with appropriate hardware and software. In some cases, even these "closed shops" may be willing to provide some level of public access to depository datasets where access to the files is important to their own teaching or research mission. For example, at UC Berkeley's Lawrence Berkeley Lab, they have mounted many library datafiles received through depository distribution on their CD-ROM network, allowing some public access to the campus community, because they considered the files important to their own research.

Clearly, if a campus computing facility is currently providing the type of in-depth support that Jim characterizes as data retrieval and analysis service to a library's primary clientele, the library would be wise to establish an arrangement to make referrals to that service rather than try to develop such capabilities in-house. Such services are so costly and labor intensive that most libraries would make better use of their resources in other areas.

Computing Support Groups

Another important source for informal collaboration and communications in support of computing services are computer users groups. Many areas, and even some campuses, have grassroots "user groups" where computer users can share information, expertise, and mentor less sophisticated users. These groups may be organized around computing platforms (IBM, Mac, etc.), software packages, or specific tasks (network administration), etc. They may meet for informal discussion, organize training sessions, or sponsor local (or even national) experts as speakers. Some may have online mailing lists. In addition, there are news groups and list servers on the Internet that deal with technical issues of interest to data users and providers. Again, these may be dedicated to a specific type of hardware or software, aspect of computing support, or substantive data issue. These groups can be invaluable in troubleshooting specific problems.

Data Libraries

Libraries should be aware of all data libraries that exist on their campus or in their local area. These may be found within computing centers, academic departments or schools, and research units. On-campus likely places to support such libraries include centralized computing centers, teaching department such economics, political science, psychology, and geography, college-level computing facilities in the social sciences and health related fields, and research units concerned with

quantitative or survey research. There may be multiple narrow subject-oriented collections in various locations on campus. Off-campus data libraries may be found in other academic institutions, city or regional planning agencies, business libraries, or research organizations. As with other computing services, they may be publicly accessible or be "closed shops" with a specific clientele. These libraries may be formally staffed and structured or run by staff or students with other primary responsibilities. If the library is interested in providing what Jim characterizes as "the lowest possible level of service," passive referral services, staff will need to be aware of the existence, holdings, and accessibility of these collections. Informal collaboration and communication will also strengthen library services as staff draw on the (sometimes substantial) discipline specific expertise in these facilities. One other possible form of cooperation is for the library to include the data library's holding in the campus OPAC.

Subject Experts

Another important source of informal collaboration are subject experts in datadependent disciplines. Library personnel will benefit immeasurably from contact with these data users. In most institutions they will tend to be members of the faculty engaged in quantitative teaching or research in disciplines such as statistics, economics, political science, sociology, psychology, management, organizational studies, public health, civil engineering, agricultural economics, education, history, anthropology, etc. Others may be in these same disciplines in post doctorate or research appointments. Many will or should be users of the library data collections. While most collaboration will be informal, this group will be the constituency best qualified to assist users with areas such as advanced datafile recommendation and datafile use advisory services. When users have advanced questions as to the content of a particular datafile and its suitability for a specific research application, or seek advice on specific research methodologies, statistical techniques or software, referrals to other more expert users in their department or subject discipline may be the only means of providing assistance. While most experts would be unwilling to enter into a formal agreement to provide public consulting on such matters, many would consider it professional courtesy to provide minimal assistance to a colleague.

Statistics Labs

Libraries in institutions with statistics labs (or the equivalent) may wish to develop cooperative relationships with these facilities. The discipline of statistics influences the method of inquiry in almost every discipline from agriculture and engineering to social and medical sciences. Campuses sometimes provide centralized laboratories in support of teaching

and research involving statistical methods. These facilities may incorporate computing equipment and range of both specialized and general purpose statistical software, as well as consulting on statistical methods and research design. Any library considering the provision of data analysis or advisory services will want to investigate the existence of such facilities on its campus. Again, collaboration may be informal and the statistics facility may only serve as a referral point for more complex methodological questions.

Local Contacts

Options may also exist for inter-institutional cooperation. Many data producers have their own distribution networks. Local members of those networks can be of great assistance and may have access to datafiles outside the library's holdings. Libraries will benefit from knowledge of and contact with any such contacts in their local area or region. Relevant networks include the State Census Data Center network, the Business and Industry Data Center network (both part of U.S. Bureau of the Census), the BEA's Regional Economic Measurement Users Group, and data centers receiving files on deposit from the National Center for Health Statistics Data Tape Program.

Another obvious option for inter-institutional cooperation is other local libraries with machine-readable collections. Cooperative support of service for datafiles may take many forms, including sharing of expertise and coordinating referrals between institutions. For example, a public library with limited data holdings is likely to benefit immensely by communication and collaboration with a larger academic institution nearby that has more extensive resources for its data services. Conversely, the large academic depository will benefit from close ties to a local public collection where the general public may be referred for basic assistance. More creative arrangements might include coordinated collection development and selection of datafiles within various subject disciplines.

Conclusion

This list is not exhaustive. It is meant to be suggestive of the types of relationships a library might seek to develop and some logical places to look for partnerships. A library's options for collaboration will be varied, and one library's options will differ from another's given their differences in institutional setting, mission, and resources.

It should be clear, however, that no library is likely to be in a position to "do it all." Financial and personnel resources will be a primary limiting factor. Even if these resources were limitless (especially unlikely in the current economic climate), there will be certain roles that are inappropriate within the traditional library model. As Jim suggests, more complex data analysis services are may

fall into this category of service. Most librarians would agree that it is not their role to evaluate the reliability of print sources, or to interpret research results or statistical tables for end user. For most libraries it will then follow that even with appropriate technical or subject background some activities are rightly outside the scope of the library's public service mission. These activities may include advising on research methodology, analytical procedures, sample design, statistical techniques, as well as software selection, and result interpretation. Unless a library has access to a comprehensive data analysis service, these activities should be avoided and specifically excluded from its public service policy.

1. Paper presented at IASSIST in San Francisco, May 1994.
2. Jim Jacobs, Data Services and Collections (hand-out prepared for the IASSIST/GODORT Workshop, Public Service for Numeric Datafiles: Issues for Depository," held February, 1994 at UCLA).