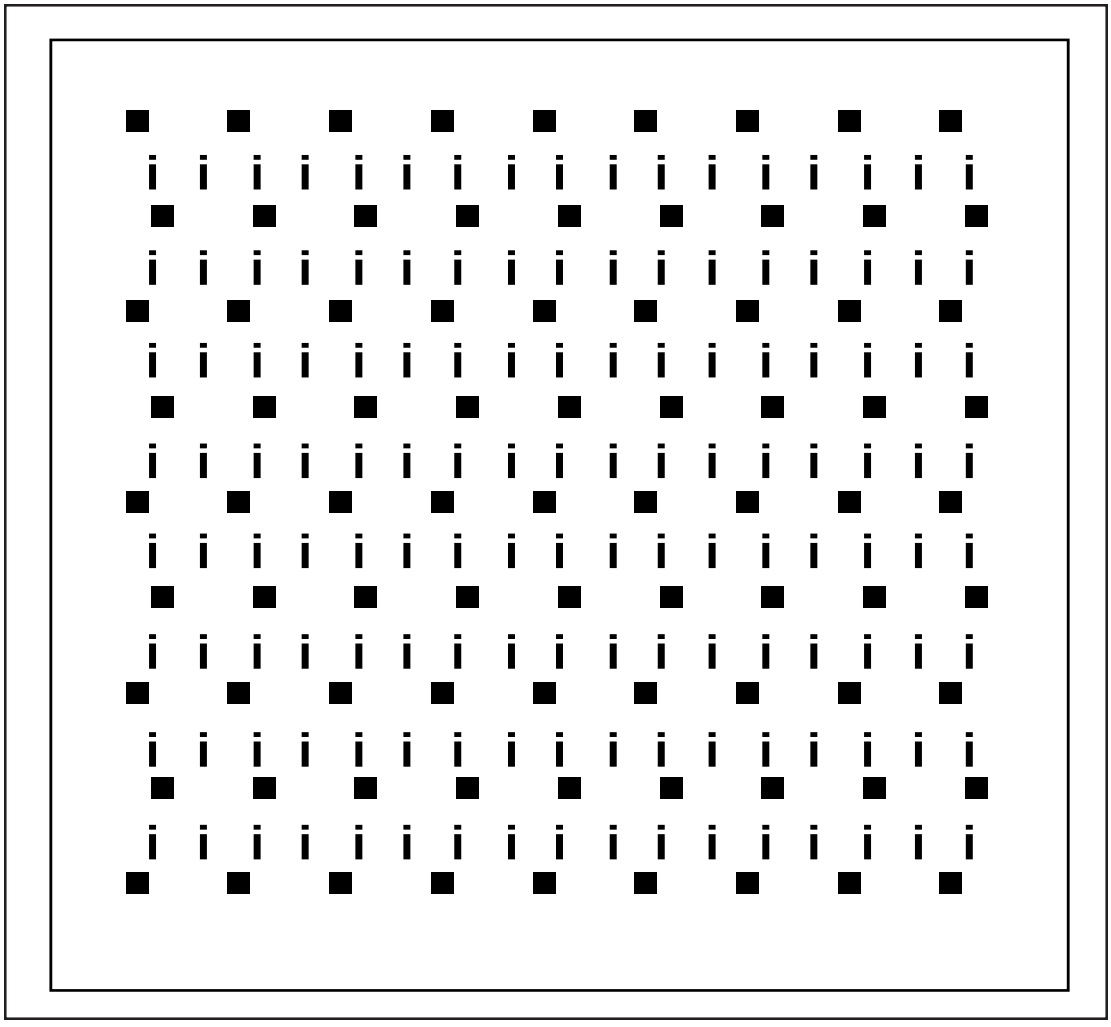
IASSIST

QUARTERLY

VOLUME 18 Fall/Winter 1994 NUMBER 3&4



Printed at UCLA

IASSIST QUARTERLY

The **IASSIST QUARTERLY** represents an international cooperative effort on the part of individuals managing, operating, or using machine-readable data archives, data libraries, and data services. The **QUARTERLY** reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine-readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of **IASSIST**.

Information for Authors

The **QUARTERLY** is published four times per year. Articles and other information should be typewritten and double-spaced. Each page of the manuscript should be numbered. The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press *Manual of Style* or Kate L. Turabian's *Manual for Writers*. Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. "Bibliographic references for numeric social science data files: suggested guidelines". *Journal of the American Society for Information Science 30(2):77-82, March 1979*. If the contribution is an announcement of a conference, training session, or the like, the text should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event. Book notices and reviews should not exceed two double-spaced pages. Deadlines for submitting articles are six weeks before publication. Manuscripts should be sent in duplicate to the Editor: Walter Piovesan, Research Data Library, W.A.C. Bennett Library, Simon Fraser University, Burnaby, B.C., V5A 1S6 CANADA. (604) 291-5869 E-Mail: walter@sfu.ca
Book reviews should be submitted in duplicate to the Book Review Editor: Daniel Tsang, Main Library, University of California P.O. Box 19557, Irvine, California 92713 USA. (714) 856-4978 E-Mail: DTSANG@ORION.CF.UCI.EDU

Title: **Newsletter - International Association for Social Science Information Service and Technology**
ISSN - United States: 0739-1137

CONTENTS

Volume 18 **Number 3/4** **Fall/Winter 1994**

FEATURES

News

Utilizing Mainframe Data on PC Platforms: Problems, Solutions, and Techniques

by Carol Wickenkamp[1]
WAE

Introduction
As more organizations and institutions downsize computer facilities in order to make greater use of the ubiquitous and inexpensive desktop computer, the problem of how to get non-ASCII data from there to here becomes increasingly common and pressing.

Archival requirements as well as data utilization are affected by the platform shift; additionally, users are expecting greater access to data than in past decades and devising access methods with and without the blessing of the MIS staff. Life expectancy of archival tape media from the 70's and 80's is diminishing. All of these issues draw us to ask the question: How do you get the data off the mainframe and onto the computer.

Let us break the big problem, the great need, into smaller and more manageable problems, in the spirit of the eating of the elephant.

Problem: Determining whether the data is even suitable for conversion
Careful evaluation of the data will help you determine whether to shelve the project or to move onward. This information is critical in determining not only feasibility, but potential cost of the project. This evaluation will help you to discover those unpleasant exceptions to the rule that will require expensive programming and special processing that can drive costs for conversion out of the feasibility range.

Techniques:
1)Check the physical condition of the media itself, especially if has been many years since cleaning and copying of the tape

2)Try to evaluate the adequacy of documentation, so far as record format, field definitions and descriptions, and code tables.

3)Do your best to determine that this data is really what you thought it would be, that it is suitable for your needs, or that isn't already duplicated elsewhere in a more accessible format.

4)Determine the tape density on older tapes, and for very old tapes, whether they are 7 track or 9 track[2].

5)Non-EBCDIC data formats crop up on older tapes especially, and can greatly increase the effort and expense of conversion. Look for packed decimal, zoned decimal, packed bit, or binary data, these data formats will need special conversion techniques[3].

6) Unusual file formats will also need special conversion techniques: for example, tapes from military sources may be in NIPS, those from medical facilities may be in MUMPS, and PICK systems have been in wide use for many years[4].

Solutions:
If your facility has mainframe to PC connections, your tapes are in good condition, and your tapes are readable by your current mainframe or mini facility, you can run a sample of 3 to 5 megabytes from each file across the network. The PC interface cards necessary for the PC to mainframe connection will automatically convert mainframe EBCDIC data to ASCII data. This sample data will help you to determine the adequacy of the available documentation and the presence of unusual data and file formats, which we will discuss further in this section. Data which does not convert directly from EBCDIC to ASCII can be readily identified. Even in very large files, a sample of this size will almost always yield usable data in all fields.

If, however, your data is truly historic, Just accomplishing this task can be a problem in itself. Unless your MIS staff is familiar with older computers, tape, and data formats, this evaluation may be better left to professionals. The section on Data Conversion Service Bureaus addresses the issue of older tapes.

Data Conversion Service Bureaus

There are data conversion service bureaus in most cities that deal with old data on a regular basis. For very old and fragile tapes, consider contacting a disaster recovery service; many of these agencies have the techniques and equipment to do serious data recovery. Be prepared to pay for this initial evaluation, and ask for a quote (based on the number of files you'll want evaluated). You will

need an evaluation that will cover all the points discussed above, in 1 through 6. In addition to the initial evaluation, request a quote for providing a 3 to 5 megabytes sample EBCDIC to ASCII conversion from each file if the tapes are readable. Unless the files are under 20 to 30 mb, ask if they can take two small samplings (500K), one from the middle of the file and one from the end of the file as part of the 3 to 5 mb sample, and find out how much extra it will cost you for these small samples. Current price for EBCDIC to ASCII conversion is usually about $10 per megabyte. Get cost quotes for your evaluations from more than one agency, and also ask if you can contact previous customers, as you would for any contract service.

Tape Drive Peripherals

If you have neither mainframe to PC capabilities nor the funding for service bureau work, or for other reasons have decided to tackle the project in-house, consider rental or purchase of a 9 track tape drive that will interface with a PC. These tape drives will come with software that will perform simple EBCDIC to ASCII conversions, and some will have software will have software with even more capabilities. For example, Qualstor's drives come with software that will convert directly from EBCDIC to Dbase. Drives are available that will handle varying tape densities; Overland makes a tape drive that will handle even the very old 800 bpi density as well as the contemporary 6250 tapes. If you know that your tapes are not fragile and you can safely run them, you can use a tape drive peripheral to do your initial evaluation of your data, running the same 3 to 5 mb sample. Data conversion service bureaus often rent drives, as do some of the larger computer equipment rental companies. The cost is usually about one teeth the purchase price; drives adequate for most conversion jobs will rent for around $600 per month.

Documentation and Identifying Unusual Formats

Using the documentation you've gathered, and a print out of your sampleASCII data (start with just a few records), you can begin the task of reading the raw data. This process will uncover gaps in your documentation as well as "funny" data. Frequently unusual data and File formats will be easily discovered on initial examination, before you even begin to check your data against the documentation. See Figure I for "Funny Data"; the fields that contain the curly brackets signal the presence of zoned decimal numeric fields, as do "/" characters and unexpected periods. Zoned decimal will be converted incorrectly in a simple EBCDIC to ASCII conversion, as is obvious.

Other non-EBCDIC numeric formats can also yield exotic results.

If you find no indication of problem data, use the field descriptions in your documentation to mark off the Fields in your data, as in Figure 3. Check your data fields one by one against both the field definitions and code table, if some of the data is coded. Here in Figure 3 we have clean data, with names, dates and Julian dates, cities, etc. where they should be and in the proper format.

Make sure that the code values in coded fields are represented in the code tables. Should you find codes that are not listed in the code table, but the rest of the data is clean and in agreement, you have probably encountered either an undocumented code (if there are many occurrences) or data entry errors. If you have undocumented codes, you can sometimes extrapolate the meaning from the data when the entire file is converted. Often a further search for more documentation is necessary. (Both the National Archives and NTIS retain copies of some Federal computer documentation.) Lack of sufficient documentation can doom your conversion project, unless you can be satisfied with either converting the portions of the data that you can identify, or just archiving the data in the hope that you can obtain the requisite documentation at a later date.

Take samples of 20 to 50 records from different places in your 3 to 5 mb sample and verify the data. If your are able to obtain records from the middle and end of your life, be sure to check them, as sometimes another file with a different format was appended to the first data file. Should you find evidence of multiple files, you will want to make a note of it so that when you have the tape converted, the data can be run off in separate files during the conversion process.

Determining Conversion Costs

Using the evaluation information about your files, you can begin to calculate costs. For example, if you send 300 mb of clean EBCDIC data to a data conversion service, and they charge $10 per mb, your charges will be $3000. To this you must add the cost of target media sufficient to store that volume of data. This figure will of course vary according to the media. Should your facility plan to download the data from a mainframe to a PC, your in-house costs will, at a minimum, include target media costs and computer time, which may or may not include computer operator charges. Coordination with your MIS department will be essential in defining costs for in-house conversion. If you have data that requires special processing, costs may include data recovery fees for very old and fragile tapes, or programming costs to convert data that is in non-standard data or file formats. You will need to obtain a second round of quotes for this

work, which will be more expensive than standard conversions, or negotiate with your MIS department for programmers to do the work. Doing the conversion yourself, for those without mainframe connections or funding for service bureau work, will be addressed in section Converting Data on a Low Budget.

If your data will require the programming services, expect to pay a minimum of $50 per hour. Programming costs in major metropolitan areas will be greater. As with other contract work, obtain more than one quote and ask to speak with previous customers. Try to speak with customers whose programming and conversion needs were similar to yours, in order to ascertain that the programmers have actually dealt with this type of data or file format; you don't want to pay for the programmer's learning curve.

Problem: Converting data on a low budget
There are those facilities who will not have the resources of an eager to help MIS department, or the budget to cover thousands of dollars for data conversion services. There are alternatives that can put the data conversion and migration process in the realm of the possible for even the most underfunded facility.

Techniques:
Hardware
Before we begin the "hands on" process of converting this data, we must have some repository for the finished product. Depending on the volume of data, there are a number of target media that will be appropriate.

High capacity hard drives are becoming very affordable, with prices dropping to around $1 per mb and even less for very high capacity drives of over 1 gigabyte. This drop in price put desktop mass storage within the reach of low budget facilities.

The lowest cost storage media will be the inexpensive PC backup tape. QIC 80 tape, which is becoming a standard for entry level backup, will store 250 mb of compressed data; this means that you will usually be able to store more than 250 mb of data on one tape. The drives are inexpensive, currently selling for under $200, and will function very well in older AT class PCs. The media will cost about $20 per cartridge. The drives are adequate for short term archival storage (not recommended for a permanent solution), but are slow and inefficient if you plan to use the data frequently.

Removable media hard disk drives are available in either internal models or portable models that interface with the PC through the parallel (printer) port; these drives offer another attractive alternative. Prices on these drives rapidly dropping; at the present, a drive in the 110-120 mb range can be purchased (with some judicious shopping) for about $400, including one cartridge; higher capacity drives are available. Each cartridge contains a hard disk platter, and the user can easily switch cartridges. The media costs are about $65, and prices should fall rapidly. The advantages include very fast access to data for those who need frequent access and portability. These drives can be compressed with disk compression utilities, increasing the storage potential. They are an excellent choice if your data files are in the appropriate size range and you will require frequent access to the data.

DAT backup drives are more expensive starting at about $1000, but they are very fast, they store gigabytes of data, and the cartridges cost about half as much as the QIC80 cartridges.

Solutions:
Data Copy by Data Conversion Service Bureau
Service bureaus will make an exact copy of your data and write it to your media. The current cost for this service will be in the range of $1 to $1.50 per mb of data. For example, if you are using QIC80 tape, request the bureau to make a copy of the data file(s) onto QIC80 cartridge media, which you will then restore to a hard drive at your facility for do-it-yourself data conversion, or simply retain as archival storage. (A discussion of do-it-yourself data conversion will follow in this section.)

If you are using a tape backup medium, be sure to tell the service bureau the name brand of your tape drive, as cartridges written by one brand of tape backup equipment be readable by equipment manufactured by another company. It is wise to do a test run with a trial tape cartridge written by their equipment, to determine whether your equipment will read the tape. You will also want to request that the data file tape headers (preliminary system information written when the tape file was created) be stripped from the data, and that only data be copied onto your medium. If you have a large number of tapes, It will be wise to pre-determine a meaningful data file naming scheme, so that you will know which data file is which when you get them back.

Nine Track Tape Drive Rental

Your facility may decide that tape drive rental is the most feasible course. Basics on PC peripheral 9 track drives were covered in an earlier topic. The company that rents you the tape drive may provide both installation and removal of the interface card if you have no one on site who can do it. As was earlier discussed, the software that comes with these drives will provide the option of converting the EBCDIC data to ASCII as it is copied off the tape and onto your storage medium. Those who are

not familiar with tape conventions such as blocking, and fixed and variable length records, determine the degree of customer support available from the rental agency You may need some initial instruction. If you have no special conversion needs, this is a most cost effective solution to the data conversion.

Data Conversion Software

Service bureaus that do data conversion and rent 9 track tape drives often sell special data conversion software that has more features than the software that is bundled with their tape drives. Typically, software of this type will handle the unusual data formats mentioned earlier, and can convert standard variable length records to fixed length records. Expect to pay $200 and up for this software. Do not count on conversion software to accomplish the task of converting the non-standard file formats discussed earlier; you probably will still require programming services.

Frequently the software interface is intimidating and may be hard to get used to, but the conversion process itself is not overwhelming. Generally, you will be required to mark off the data fields (as you did with your sample, only on screen rather than on paper) and then define the conversion process that is to take place, i.e.. EBCDIC to ASCII, binary to ASCII, or packed decimal to ASCII. When you have defined your conversion instructions, your file is ready to be converted by the software.

It is a good idea to run a partial conversion of 500 to 1,000 records to verify the accuracy of your field definitions. Sometimes the process will require several tries before all the bugs are out of your conversion instructions, and it is far faster to convert 1,000 records for a sample than to convert 100,000 records. The speed of conversion will depend upon the processor speed of your computer, the complexity of your conversion instructions, and the length of your records. You can use the measure of 1 megabyte per minute as a rough rule of thumb. Although most of these programs will operate on files residing either on the tape drive or a hard disk, it is much faster to copy your file onto a hard disk and do the conversion from disk.

Problem: The data is so heavily coded that it will be difficult to work with

As a rule, database programming relies heavily on code table to hold frequently used values; old mainframe data can be coded in every field, thus yielding very compact files. The code values were replaced at processing time so that reports were understandable. This sort of data is very cumbersome to use, even with modern and easy to use database programs such as Paradox, Alpha Four, Access, etc.

Solution:

Given the low cost of hard disk storage, It is becoming more feasible to simply replace the coded fields in databases with their values, yielding a significantly larger, but easier to use flat file database. Even with a two or three fold increase in file size, this solution can bring comprehensible, easy to manipulate data to the most unsophisticated user. It is far faster and more accurate to extract reports or meaningful data screens from a database that contains “Lutheran” rather than “07”, “Buick” rather than “15” or “CA” rather than “05”.

Expert programming skills are not necessary to accomplish these replacements, a moderately skilled in-house programmer should be able to do the job. Even if it is necessary to hire a programmer, it should not be a major expense, unless you have a large number of heavily coded files.

Conclusion

Although moving data from older mainframe generated tapes to a PC platform is a process that requires planning and attention to detail, the task is not insurmountable, nor is it always exceedingly expensive. With the exception of very old or non-standard tapes, much of the work can be done in-house and with a small budget, utilizing moderate computer skills.

Notes:

2. 7 track is an obsolete tape standard which used the 6 bit BCD (Binary Coded Decimal) code together with a parity bit. The contemporary 9 track drives will not read 7 track tapes.

3. Although data conversion software renders these numeric formats harmless to the non-technical user, a discussion of these formats is included for those who are interested. Numeric data format which will not convert in a standard EBCDIC to ASCII conversion include:

Packed Decimal with low order sign bit

This is the normal IBM packed decimal field.

Zoned Decimal with low order sign bit

This format is generated by some COBOL, PL/I and Assembler systems; although not common, it is still in use in some contemporary installations. Zoned Decimal is a standard EBCDIC numeric character field with the exception of a sign code in the high order nibble of the low order byte, with C hex and F hex being a positive sign code and D hex a negative sign code. This results in invalid EBCDIC characters in the low order byte of some zoned decimal fields.

Binary with most significant byte first

This is the format in which IBM mainframes normally

process binary data; normal PC binary format is binary with least significant byte first

Packed with high order sign bit

This is a binary format with the sign bit in the high order nibble of the high order byte.

Packed with no sign bit

This is a normal packed field, except that all nibbles contain a significant digit (no sign field) and the field may begin and/or end on a nibble boundary.

4.These are all non-standard variable length file formats. MUMPS has been widely used in VA hospitals and in medical clinics, and is still common. PICK usage extends across the commercial spectrum. NIPS was designed specifically for use on IBM 360 computers, and is no longer in use.

Sources:

* For further information on tape formats, labeling and file conventions, you can contact:

American National Standards Institute, Inc.
1430 Broadway, New York, NY 10018.
Tel : (212)6424900

Ask for publication X3.27, “Magnetic Tape Labels and File Structures”

IBM tape labeling conventions are explained in the IBM publication “OS/VS Tape Labels” (GC26-3795-3, File No. S370-30) and “DOS/VSE Tape Labels” (GC33-5374-1).

DEC information is described in “Guide to VMS Files and Devices” (AA-LA06A-TE), available from DEC.

* If your facility is not in a metropolitan area, you may find several reputable data conversion service bureaus advertised in PC Magazine, which is available in most drug stores and supermarkets.

* Two companies that produce data conversion software, each with different capabilities, are:

NovaStor
30961 Aguora Road, Suite 109
Westlake Village, CA 91361
(818)707-9900
Fax (818)707-9902

Overland Data
5600 Kearny Mesa Road
San Diego, CA 92111
(619)571-5555
Fax (619)571-0982

Service bureaus may also have information on other data conversion software.

1. Paper presented at IASSIST 1994 in San Francisco.. Reprints of this paper are available from: Carol Wickenkamp, WAE, PO Box 349, Clarkston, WA 99403

For Better or For Worse: academic partnerships for data services

by Diane Geraci[1]
Binghamton University
State University of New York

Introduction
While there is no one model for providing services for data in colleges and universities, it is increasingly common for various constituencies to cooperate, especially in lean fiscal years. There are both positive and negative aspects to pooling resources in such a "marriage of convenience." Although not the solution for everyone, this paper will take a look at a partnership among two academic departments, Computing Services, the Libraries, and the Provost's office at Binghamton University, State University of New York. ' It will suggest advantages and disadvantages for those considering cooperative ventures at their institutions.

From this day forward,
for better for worse, for richer for poorer,
in sickness and in health,
to love and to cherish,
till death us do part ...

Data library and service operations in academic institutions in North America have in many instances seen a reduction in resources in the last five years. In some cases, this has threatened the existence of some or even all services. In others, it has caused data professionals and administrators to be creative and forge new arrangements to maintain or even enhance basic levels of service for their clientele. Because data service organizations vary considerably from academic institution to institution, there is no single or simple way to diagram a preferred organizational structure for data service. What works in one academic setting, may not in another. Services seen as basic at one university may be on a wish list at others. Size and diversity of user groups also vary depending on programmatic and research agendas. In any case, optimal staffing and funding levels are directly related to the level of service needed by an institution's primary clientele. Unfortunately, even minimal resource levels may not be possible at some institutions.

Pooling resources among departments and units across a college or university can be an option where a separately funded "data center" or "data library" does not exist, or, when an existing service is faced with dissolution. While these 'marriages of convenience not suitable for all organizations, there are significant advantages and disadvantages of academic partnerships. They are especially worth exploring if an institution faces "rightsizing" or consolidating services. These partnerships rely on the ability of various constituencies to work together, an agreed upon common purpose, mutual respect, and tolerance.

from this day forward...
An institution's history of providing quantitative, social research support on a given campus will often set the stage for future service configurations. Because of this it can be difficult to change support paradigms, although it is certainly possible and even necessary in some cases.

At Binghamton University, State University of New York, the Political Science Department in conjunction with an organized research center, provided support for quantitative social data for two decades. In 1990, a time of considerable fiscal uncertainly in the University system, the impending closing of that research center necessitated rethinking the way in which we were organized to provide data services. For the most part this meant fulfilling our Inter-university Consortium for Political and Social Research (ICPSR) membership responsibilities and related data services.

After a series of extensive consultations with administrators, faculty and staff, the Libraries agreed to assume responsibility for 'data services.' This primarily entailed maintaining formal relationships with ICPSR and later the U.S. Bureau of the Census' State Data Center Program. Ultimately this meant that the Libraries would:

. maintain formal relations with ICPSR

.serve as liaison for the State Data Center Program

. assume fiscal responsibility for ICPSR membership after an initial transfer of monies from the Provost's office

. provide customer services, particularly identifying and ordering data

.collect and maintain codebooks, related technical documentation and statistical manuals

.provide user consultations, research assistance and referrals .cooperate with Academic Computing, to make data available and to provide complementary services

.cooperate with the Economics and Political Science departments, and the Assistant Provost for Graduate Studies and Teaching to assign two ICPSR/Data Services graduate assistants to the Libraries.

The formal change in service occurred in July 1 1991 to coincide with the new fiscal year. However, Academic Computing, the Libraries, the Political Science department, and the organized research center had already begun the process of working together several years before. This early period effectively served as a 'getting to know you" phase where each unit's service orientation and working patterns became known. Evolving service plans and position descriptions assisted in making clear who would be responsible for which aspect of the reconstituted service.

for better for worse...
Commitment of each constituency is essential for a service that exists through the shared agreement of its partners. The best strategy for success is creating a win-win situation whereby each of the partners benefits from contributing to the service. A benefit may mean better meeting the mission of the unit, such as a library or computing service that serves the entire academic community. From an institutional point of view it may mean reducing duplicate purchases or services. It certainly will mean providing the kind of research support desired by relevant academic programs. It can also mean acknowledging that going it alone might not provide the depth and range of services needed.

While good will and intentions may characterize a shared agreement to provide service, a written plan is well worth the effort. Support staff and administrators do change. A written service plan cannot absolutely guarantee the continued cooperation of each unit but it does provide a framework and codification of responsibilities.

After seven years of sharing responsibility for data services on the Binghamton campus, several benefits are evident. They include:

• ICPSR membership benefits are more widely available to all constituencies on campus. There had been a perception that everyone knew about the ICPSR and the extent of their data holdings. This turned out not to the case. New faculty and graduate students continually arrive on campus and existing campus instructors and researchers have new data needs. Researchers in departments not traditionally employing quantitative research methodologies may begin doing so. There is a continual need for dissemination of information about new data and related data news. For example, only one department knew about the ICPSR Summer Program in Quantitative Methods before the Libraries coordinated the membership services.

• Duplication of data acquisitions was reduced. Because data support originally resided in the school of arts and sciences, other schools and divisions often bought there own data directly from producers. We found that much of these data were available via our ICPSR membership. This was especially true for health data and economic time series data.

• Integration of data collected in several media is a positive by-product of centering access to data in the Libraries. Print resources, CD-ROMS, diskettes, remote access via the Internet, and commercial services already are available in or through the Libraries. Making the Libraries the first stop to ascertain if data are available on mainframe cartridge tape has brought together conceptually if not physically, access to related resources.

• Existing expertise is utilized; that is, information management skills, computing skills and service orientation in the Libraries; technical, computing and statistical skills from Computer Services; research skills of the departmental graduate assistants.

• Skills shared across units increase the skills of all contributors to the service. Graduate students particularly gain solid experience working with data and valuable statistical programming skills.

. Cooperation with other units on campus increases awareness of research needs as well as understanding of different campus cultures. Daily contact with colleagues in other campus units greatly fosters understanding and respect for each other's work.

Several difficulties or less positive aspects of the partnership also became apparent.

We also found:

• ICPSR resources became more widely used on campus making it difficult for part-time staff in the several units providing support to keep up with demand. Statistics showed a substantial increase in data use on our campus as a result of the reconstituted

service. Staff in the Libraries and in Academic Computing found that an increased percentage of their work week supported data services. Some reorganization of duties occurred in each unit with the pressure being born by existing staff members. Similarly, the service began with one graduate assistant. It soon became clear that one was insufficient and we were able to negotiate for another student.

- Reliance on graduate student support entails constant training and rotation of staff. Considerable fluctuations in the quality of service regularly occur.

- Additional permanent staff is desirable, but thus far, has been unattainable. Research level support is very time-consuming. Permanent staff and new lines are difficult to acquire. They would assist in providing consistent service and allow for performance of needed tasks, especially as the number of users increases and users' request an increased level of service.

- Keeping current with data services developments requires additional space and equipment. Changes in computing platforms and storage devices require new hardware and software. Decisions made in one unit may affect another. For example, the decision by Computing Services to stop maintenance of 9-track tape drives has consequences for the way that the Libraries order data.

- New skills are required. For example, knowledge of database maintenance, cataloging, or statistical programming, and understanding research design may be necessary for data services staff to provide certain services. For already overextended staff, there is not adequate time for learning new processes or acquiring necessary skills. The aptitudes of existing staff for acquiring new skills will also vary.

- Cooperating with other units on campus is difficult in practice. Conflicting priorities in a unit or between units may be difficult to resolve. Politics internal to a unit are less easily negotiated by those outside the unit. Service orientations or philosophies of the partners may differ.

in sickness and in poor-health...

In times of staff reduction, fiscal uncertainty, competing demands in a unit, or simply a reprioritization of needs or goals, a joint service can suffer the consequences. There can be real concerns for the integrity of the service as a whole if a key group withdraws its support. When individual units experience shifting priorities or staff reductions the danger exists that the shared service will fall to the bottom of the list of things to do, or worse, will no longer be supported. When there are administrative changes the partners in the service may need to renew their "vows."

While living with a small degree of uncertainty is admissible, a crisis can arise if one contributor to the service can no longer participate or even temporarily suspends participation. Major disruption of service or stress on the other partners can occur if one unit is unable to meet their obligations.

There is not a way to absolutely ensure that no change will occur in a partner's commitment to the relationship. There are ways, though, to engender support for the service and keep it on the priority list of each partner. Relying on a core group of researchers as an "advisory group" is one way to get feedback from users. Measuring the amount of data ordered, number of users assisted, computer usage, and any other relevant factor at an institution can demonstrate the utility and necessity of the data service to administrators.

to love and to cherish...

When there is stability in the service and researchers' needs are being met, all partners deserve congratulations for cooperating across units and effectively working together to create a viable service. This is the 'feel good" outcome of a win-win situation and should be enjoyed. Lest complacency cause problems, it is a good idea to reaffirm what works with the arrangement and what can be handled in a better way. Assessment

during the good times is much less threatening then when the sky is falling due to impending budget cuts or some other "natural" academic disaster. Several methods work well to evaluate the service including meeting with the primary front line staff in each unit, consulting an advisory group of researchers, and surveying past and prospective users of the service. Taking the time for assessment is a positive way to renew the agreement and service plan(s) of the units involved and make any necessary adjustments.

till death us do part?

Binghamton's "marriage of convenience" came at time when data support on the campus was in jeopardy. It has served the university community well in its time. It does not mean that this is the only way to provide data services or that another type of service will not evolve from it.

There are several reasons a partnership such as Binghamton's might cease to continue:

. The service is no longer necessary. There may be

other ways to meet the need of data users. Schools or departments might decide to provide some of their own services. National or international consortia and computer networks may provide more data services negating the need for some local services. It is difficult to imagine, though, that some measure of local support will not be necessary, even in a future of distributed services over "the net." There certainly will be a time when the service needs to be reformulated or reconstituted.

. One or more of the partners cannot afford the commitment of staff and/or resources. A worst case scenario is the service dies. Another possibility is that the other partners are able to pick up the slack. In the case where the partners are unable to absorb additional responsibilities, providing a reduced level of service may be necessary.

. Cooperation is no longer possible between the partners. One of more of the partners may experience a change in their mission, unresolvable disagreements may occur between partners, or administrative prerogative may preclude further cooperation.

Providing data services through an academic partnership can be very rewarding. Forging key relationships between disparate units and seeing positive results in support of research and teaching are successful outcomes. Before embarking on a cooperative venture, careful consideration of a partnership model's suitability for the needs and culture of an institution is necessary.

1. Paper presented at IASSIST 1994 in San Francisco.

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 files[2]. 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 it 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 pubic 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 pubic 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 Intemet 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 anothers 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 Francsisco, 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).

Gopher Servers as a Point of Access

by Julie A. Fore[1]
Assistant Automation Librarian
Indiana University Ruth Lilly Medical Library

This paper will discuss why and how to use gopher servers on the internet to provide access to locally developed data. This includes formatting the data, and establishing the links to the data on the server. The responsibilities involved with providing information on the internet will also be discussed.

Reasons to Mount Data on the Internet
The internet is a vast source of information and chaos, why would anyone want to add to it?

Some possible reasons for mounting data on the internet include:

-The internet allows remote access to the data. The data and its users are not limited to physical locations. This means that people from other institutions, as well as your own, can get to your data.

-The internet allows 24 hours a day, 7 days a week access to your data (barring the usual network outages or maintenance down time for the data server).

-By providing the data on the internet, it is by definition in electronic format. This allows for further manipulation or massaging of the data. It allows users to take advantage of the computer's abilities, such as searching and sorting.

-The internet allows for quick and easy publishing or updating of your data.

Reasons Not to Mount Data on the Internet
Most of the reasons for not mounting data on the internet are based on privacy and legal issues.

-The data is copyrighted and can not be re-distributed.

-The data is of sensitive nature and should not be accessible to just anyone, i.e. the world.

-The data is already out on the internet, in a number of different places.

-The internet location for your data is unreliable or not maintained by anyone.

Why Use a Gopher Server
Currently, gopher clients are widely distributed across the internet community and are available for most types of computer hardware and operating systems. Most gopher client and gopher server software does not require high level computers on which to run, unlike other internet tools such as Mosaic. Gopher clients, as a rule, are easy to use. They provide a common interface to many different types of resources. The Gopher protocol provides the capability to perform searches on databases and files. Currently, this is mostly primitive string or character-by-character searches. Gopher servers have the ability to link or point to other Gopher servers. This linking ability makes it relatively easy to create subject oriented gopher servers.

Indiana University Ruth Lilly Medical Library Gopher Server Pilot Project

The Indiana University Ruth Lilly Medical Library has been maintaining a database of its Permanent Reserve Collection holdings using a bibliographic database management system called Pro-Cite, made by Personal Bibliographic Systems (PBS). The Library uses Pro-Cite to keep this database because the software allows the Library to provide a number of different printouts for the library patrons to use. The Reserve Collection is shelved (mainly) in title order. Pro-Cite allows the Library to generate printouts of the collection in shelflist (title) order, as well as lists sorted by author or subject heading. (fig. 1) The Library patrons make great use of these printouts, as do the Library Circulation Staff.

RESERVES Collection by AUTHOR

Abbas, Abul K.
Cellular and molecular immunology.

Abdellah, Faye G.
Patient-centered approaches to nursing.

New directions in patient-centered nursing; guidelines for systems of service, education, and research.

Ackermann, Uwe.

Essentials of human physiology.

[...]

Fig. 1 — "Reserve Collection by Author" - Printout from Pro-Cite.

The Permanent Reserve Collection database is small at about 225 records. This made it a perfect pilot for testing how well the Library's various Pro-Cite databases would make the transition from in-house use only to internet accessible information.

The pilot project started with an analysis of the data and data fields already in the Pro-Cite database. (fig 2)

Rec# 780
Auth Abbas, Abul K.//Lichtman, Andrew H.//Pober, Jordan S.
Titl Cellular and molecular immunology
PlPu Philadelphia
Publ Saunders
Date 1991
Extn xi, 417 p
ISBN 0721630324
Call QW 568 A122c 1991
Desc Cellular immunity/Immunity—Molecular aspects/Immunity, Cellular/
Lymphocytes- -immunology

Fig. 2 — Example of a bibliographic record in Pro-Cite

After the evaluation of the electronic data, a decision was made as to which data elements would be most valuable to a person accessing the database over the internet. I decided to use basic bibliographic citation fields, i.e. author, title, place of publication, publisher and date; as well as the subject heading information. The data from these six fields were then exported from Pro-Cite using Pro-Cite's import/export utilities. Pro-Cite created a standard comma (",") delimited file. (fig 3.)

"Abbas, Abul K.//Lichtman, Andrew H.//Pober, Jordan S.","Cellular and molecular immunology","Philadelphia","Saunders","1991","Cellular immunity/Immunity—Molecular aspects/Immunity,Cellular/Lymphocytes—immunology"
"Abdellah, Faye G.","Patient-centered approaches to nursing","New York","Macmillan","<1960>","Nurse-patient relations/Education, Nursing"

Fig. 3 — Sample of Pro-Cite Export File in Comma Delimited Format.

The Gopher Server software being used by the Ruth Lilly Medical Library, KA9Q NOS, requires database files to be in dBASE III or dBASE IV format. While most current database management programs such as Paradox by Borland and R:BASE by Microrim can save data in dBASE III format, we chose to use the dBASE III program for the next part of the pilot program.

A database structure was created in dBASE III using the six fields exported from the Pro-Cite database. To keep things simple, the Pro-Cite field labels were used as the field labels in the dBASE database. While Pro-Cite, for the most part, does not use fixed field lengths, dBASE requires fixed field lengths. We made educated guesstimates on what the dBASE field lengths should be. Figure 4 shows the final structure for the Reserves dBASE database.

Structure for Database: C:RESERVES.DBF
Number of Data Records: 225
Date of Last Update: 4/21/94

Field	Field Name	Type	Width	Dec
1	Auth	Character	130	
2	Titl	Character	200	
3	PlPu	Character	50	
4	Publ	Character	50	
5	Date	Character	8	
6	Desc	Character	254	
** Total **			693	

Fig. 4 — Final dBASE III File Structure

The dBASE III import function was used to convert the Pro-Cite produced comma-delimited file into a dBASE III database. A paper report of the new database was then created to verify two things. First, that the data was correctly transmitted from Pro-Cite to dBASE III. Second, to verify that the data, itself, was correct and complete. The data had indeed transferred correctly, but it was found that some of the records in the original database contained incomplete information.

Once the dBASE database had been cleaned up, and an ascii text file bibliography was generated from it using R&R Report Writer by Concentric Data Systems, the data was ready to be transferred to the actual microcomputer running the gopher server software. In the case of the Ruth Lilly Medical Library Gopher Server, this meant taking down the gopher

server, that is, exit out of the server program. Then, using the DOS copy command to move the files from the Library's Novell file server to the gopher server's DOS-based microcomputer. Once the actual files were residing on the gopher server's hard drive, a suitable access point in the gopher's menu structure had to be found. Finally, the gopher server's menu configuration files had to be edited to include the pointers to the files.

The most logical place to include the reserve collection information was in the menu with all the other files specific to the Ruth Lilly Medical Library, i.e. the files containing the Library's hours, policies, journal holdings, etc... (fig. 5)

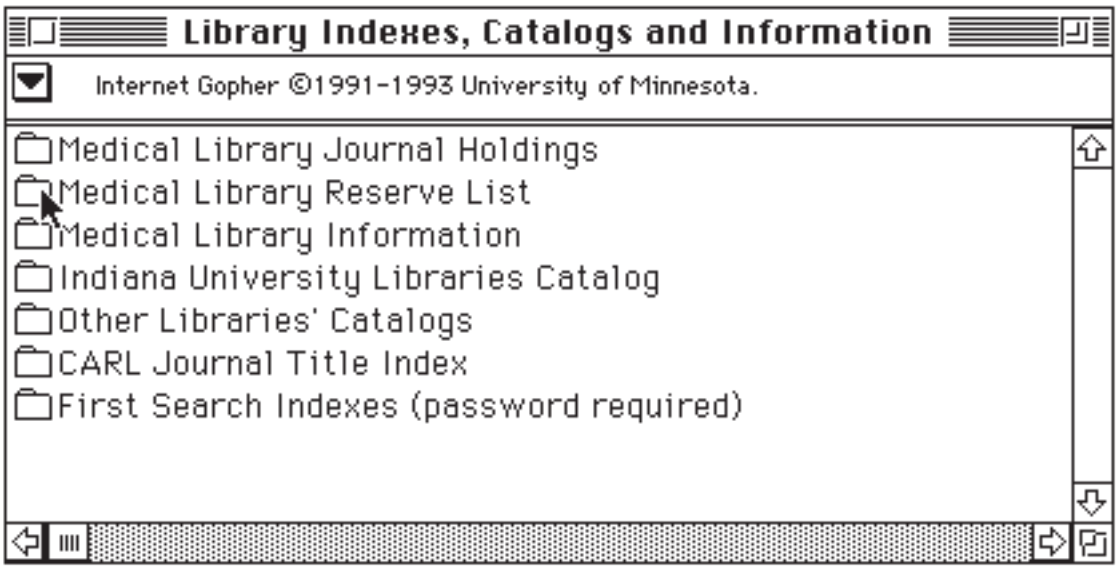


Fig. 5 — "Library Indexes, Catalogs and Information" Menu from Ruth Lilly Medical Library Gopher

In the KA9Q gopher server software, the menus are designed using directories and subdirectories on the server's hard drive and "GINFO" files (which possibly stands for "gopher information" or "gopher index file"). For every menu on the server (seen by a gopher client) there is a corresponding directory on the hard drive of the gopher server and in that directory a GINFO file. The GINFO file contains five elements: 1) the text shown on the menu to a gopher client, 2) a code for the type of resource that is being pointed to (text file, database, directory, Macintosh Binhexed file, uuencoded file, GIF file, etc...), 3) the name and path of that resource (for example /server/reserve.db/reserves.dbf), 4) the internet address of the gopher server that provides the resource (for example gopher.medlib.iupui.edu) and finally, 5) the port for that gopher server, usually port 70. An example of a GINFO file is seen in figure 6. The GINFO file is were the telnet or ftp links to other internet sites are described.

```
1Medical Library Reserve List              1c:/server/reserve.db  gopher.medlib.iupui.edu   70
1Medical Library Information               1c:/server     gopher.medlib.iupui.edu            70
1Medical Library Journal Holdings          1c:/pub   gopher.medlib.iupui.edu          70
1Indiana University Libraries Catalog      1c:/pop/catalog  gopher.medlib.iupui.edu  70
1Other Libraries' Catalogs                 1/Libraries   yaleinfo.yale.edu              7000
1CARL Journal Title Index                  1c:/library/CARL      gopher.medlib.iupui.edu   70
1First Search Indexes (password required) 1c:/library/FIRST  134.68.85.17         70
.
```

Fig. 6 GINFO File for the "Library Indexes, ..." Menu of the Ruth Lilly Medical Library Gopher Server.

Figure 7 illustrates what might be found in a directory on a gopher server, note the presence of the GINFO file.

```
 Volume in drive M is BVOL
 Directory of M:\GOPHER\LIBRARY

GINFO              523 05-13-94   8:48p
CARL       <DIR>       05-23-94  10:35a
FIRST      <DIR>       05-23-94  10:35a
      3 file(s)          523 bytes
                 191,299,584 bytes free
```

Fig. 7 — Listing of the files in the Sub-Directory containing the GINFO file from Figure 6.

The second line of the GINFO file shown in figure 6 is the pointer to the Library's Reserve Collection Menu. As one moves through the Gopher's menu structure, the Medical Library Reserve List Menu eventually appears. (fig. 8)

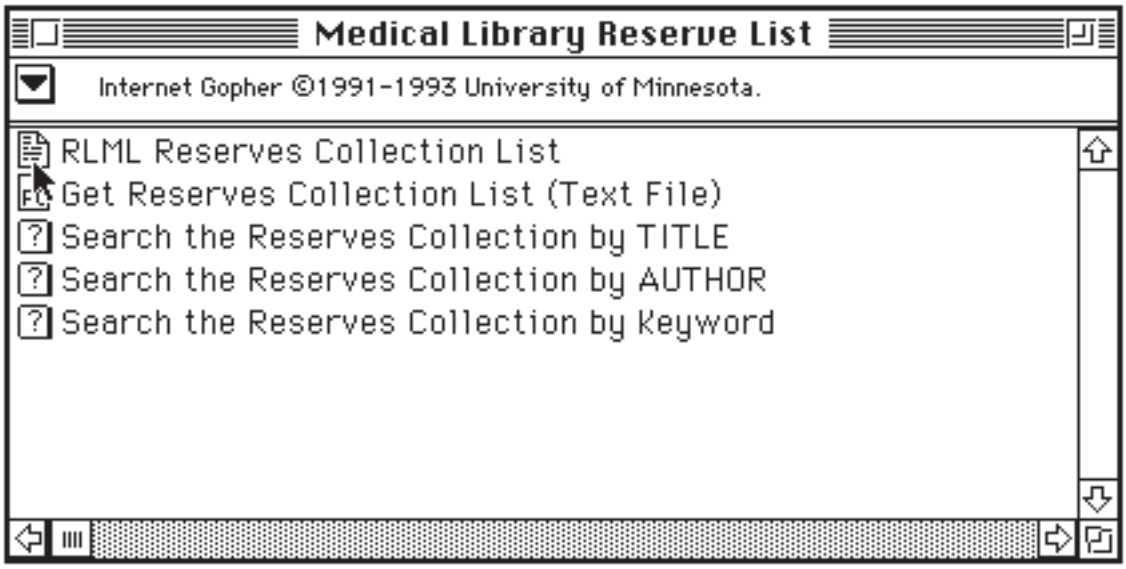


Fig. 8 — "Medical Library Reserve List" Menu of the Ruth Lilly Medical Library Gopher Server.

The "Medical Library Reserve List" menu allows a gopher client to browse or page through a text file bibliography of the Reserve Collection, ftp (file transfer protocol) the bibliography back to the user, or perform a character search on the database either using the title field, the author field, or the descriptor field. The GINFO file for this menu determines which function is performed on which file. There are only three files in the directory for this menu. The GINFO file, the actual database file called reserves.dbf, and the text file bibliography called reserves.txt. (fig. 9)

```
Volume in drive M is BVOL
Directory of M:\GOPHER\SERVER\RESERVE.DB

RESERVES DBF      156,160 04-21-94   3:46p
RESERVES TXT       30,989 04-25-94   4:17p
GINFO                 530 05-21-94   2:12p
      4 file(s)       188,210 bytes
                191,299,584 bytes free
```

Fig. 9 — Directory listing of the RESERVE.DB subdirectory

Figure 10 illustrates the GINFO file for the Medical Library Reserve List menu.

```
0RLML Reserves Collection List   0c:/server/reserve.db/reserves.txt              gopher.medlib.iupui.edu
70
5Get Reserves Collection List (Text File)          5C:/SERVER/reserve.db/reserves.txt
gopher.medlib.iupui.edu                  70
7Search the Reserves Collection by TITLE        qc:/SERVER/reserve.db/reserves.dbf~TITL
gopher.medlib.iupui.edu                  70
7Search the Reserves Collection by AUTHOR  qc:/server/reserve.db/reserves.dbf~AUTH
gopher.medlib.iupui.edu                  70
7Search the Reserves Collection by Keyword   qc:/server/reserve.db/reserves.dbf~DESC
gopher.medlib.iupui.edu                  70
.
```

Fig. 10 — GINFO for "Medical Library Reserve List" menu

Figure 11 shows what the text file bibliography looks like when viewed by a gopher client. The bibliography file was created so that the users could have access to a formatted file that they could browse through. It was decided that the dBASE database format was not very easy to browse (fig. 13), nor was it in a file format that most people could use once they had it back at their own computer. The double slash marks (//) in the author field are left over formatting codes from Pro-Cite. These codes will be removed the next time the database needs significant updating.

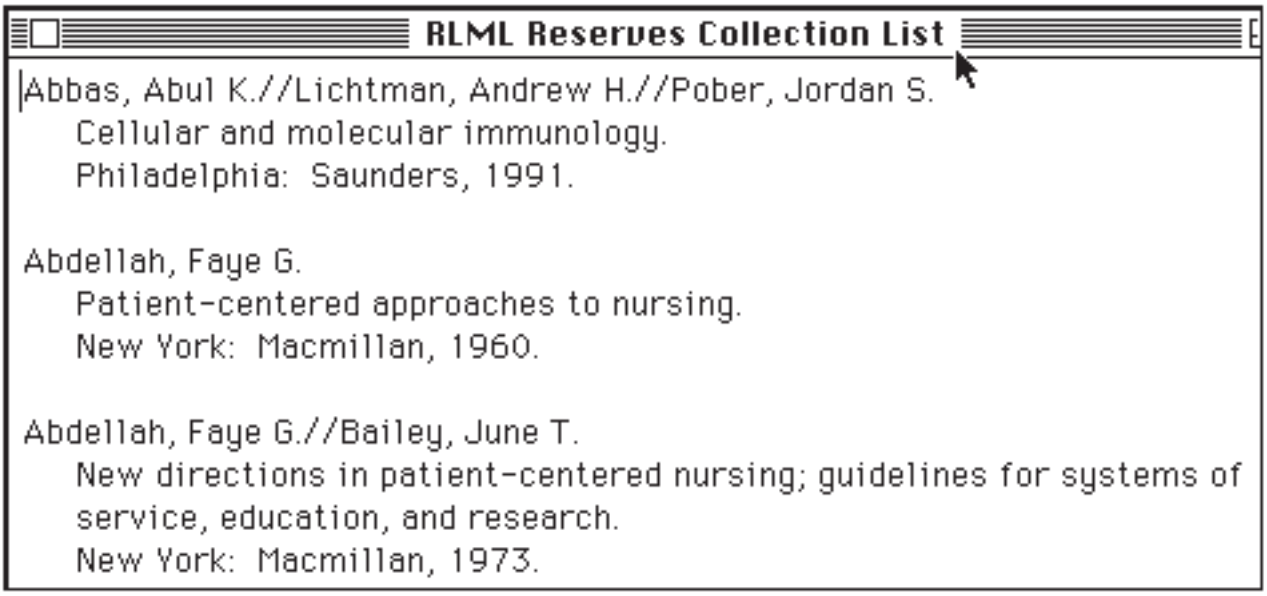


Fig. 11 — Browsing the "RLML Reserves Collection List" Option

The KA9Q gopher server search capabilities are currently string or character based searches. This means that when "Search the Reserve Collection by AUTHOR" is selected off the "Medical Library Reserve List" menu, a dialog box will appear asking the user to enter the words to be searched for in the author field of the database. In the example illustrated by figures 12 and 13, the user asked the gopher to search for all occurrences of the word "sid" in the author field. As the results of the search show (fig. 13), the gopher does not care where the "word" "sid" appears in the author field. It found the letters, or characters, "s-i-d" in the word "President" and in "Sidney". It is expected that gopher-based searching will improve in the future. If not, then some other internet tool will take gopher's place.

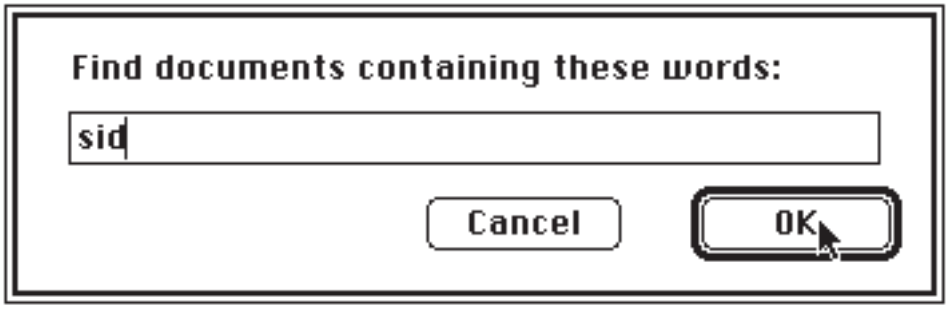


Fig. 12 — Gopher Search Dialog Box

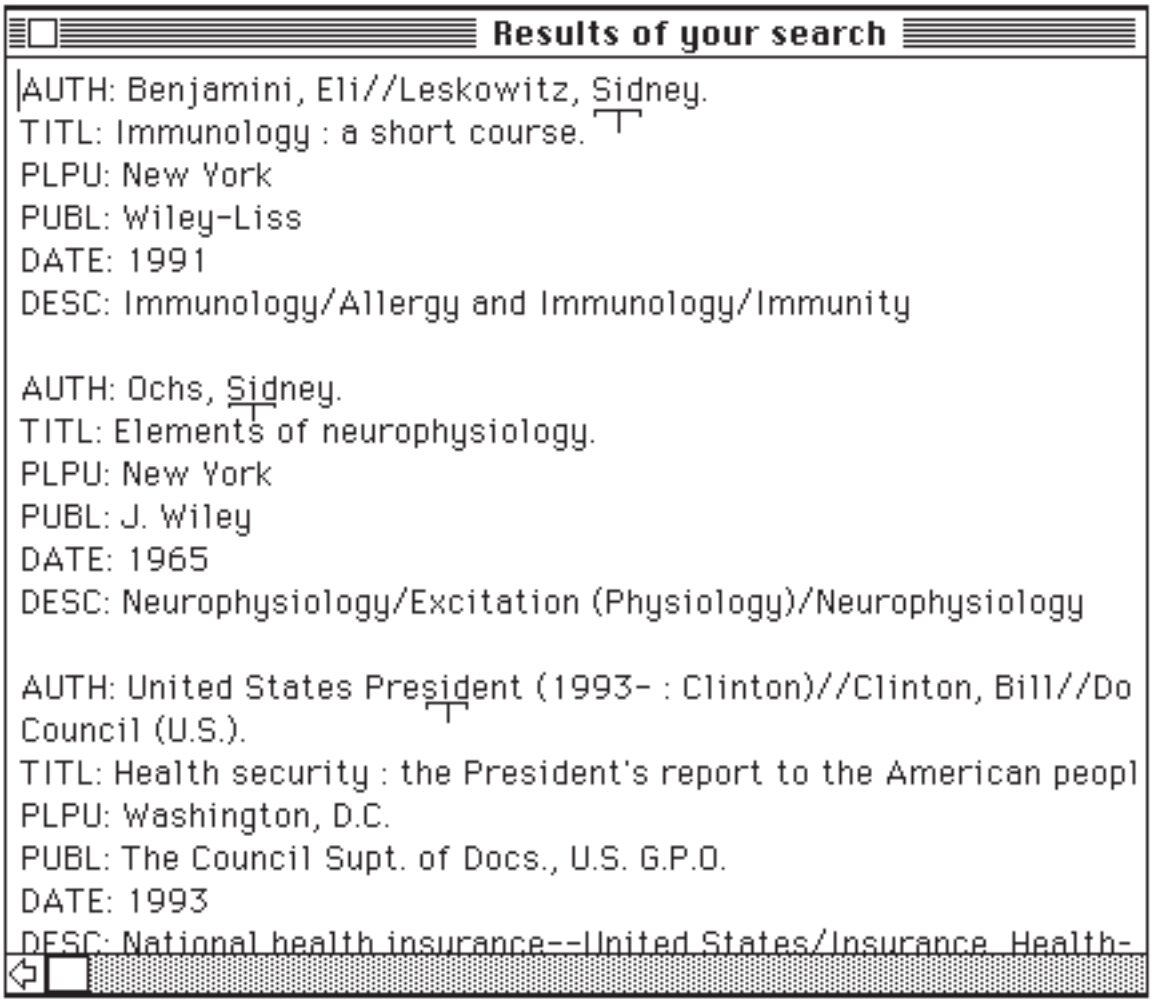


Fig. 13 — Results of Searching for "Sid" in the Author Field of the Reserve Collection Database

Internet Responsibility

It is not enough to just mount a database on the internet. It is necesary to take responsibility for it and for the gopher server on which it resides. There are a number of points to keep in mind when setting up and maintaining servers on the internet.

-Keep your server up and running. No one can use your data if your server or your network is down.

-When (not if) you take your server down for routine maintenance, i.e. on a routine schedule, post this information on your server.

-If there are limitations to your server or your data, post this on your server and on any public announcements you send out about your server. Some examples of limitations might include, access only during non-business hours like 5 p.m - 6 a.m EST, a limited number of simultaneous users, the fact that passwords are required for access to certain files or services, or that only users from a certain place (campus, university, etc...) are permitted access to a resource.

-Keep your data current and accurate. If this is not possible, indicate on the server that the data is old/out-of-date or not necessarily accurate.

-If you move your resource to a new internet site or remove it from the internet, announce this. Place a notice stating the new location of the resource in the old location of the resource. Post announcements to appropriate LISTSERVs and newsgroups.

-If you are keeping copies (mirrors) of your resource at more that one location, keep them current and announce their locations as well.

Technical Information about the Indiana University Ruth Lilly Medical Library Gopher Server
URL: gopher://gopher.medlib.iupui.edu port 70

The IU RLML Gopher Server is currently running on a Gateway 2000 386-25 MHz processor with 4 MB of RAM. The computer has a 300 MB Hard Disk, of which approximately 50 MB is being used. The computer is running DOS Version 5.0 and is attached to a 4 Mbps Token Ring LAN. The Server is backed up weekly to a Novell Netware 3.11 file server.

The Gopher Server Operating System is KA9Q NOS, a DOS-based Network Operating System. KA9Q supports Gopher; POP2, POP3, and SMTP mail server protocols; ftp, anonymous ftp, telnet and finger; CSO Name Server functions; NTP (time) Server functions; and WWW Server functions. The Indiana University Ruth Lilly Medical Library currently is not supporting the mail server functions but is experimenting with the other capabilities of the KA9Q software.

In the future, the Indiana University Ruth Lilly Medical Library Gopher Server will be switched to a 10 Mbps Ethernet LAN. It MAY be switched to a UNIX - based computer, and it may be given additional WorldWideWeb (WWW or W3) functionality and resources.

Places to find more information

Newsgroups for gophers and other information servers:

comp.infosystems.gopher
comp.infosystems.www
comp.infosystems.wais

Frequently Asked Question (FAQ):

Gopher FAQ can be retrieved via anonymous ftp from the following site:
rtfm.mit.edu:/pub/usenet/news.answers/gopher-faq
or via gopher from:
129.130.10.5 port=70, path=0/Frequently Asked Questions (FAQ)/gopher-faq

KA9Q NOS (Network Operating System) DOS-based Gopher Server Software.

KA9Q Mailing List:
send an email to Ashok

ashok@biochemistry.cwru.edu and ask to be added to the mailing list. This address is an individual, so be nice.

KA9Q Manual:
The User Manual is available via gopher from the following site:
cases.pubaf.washington.edu, port 70, in 1c:\manual

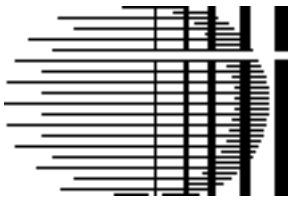
University of Minnesota — The Top Gopher:

gopher://gopher.tc.umn.edu port 70
Questions or Comments for the Gopher development team, send e-mail to:
gopher@boombox.micro.umn.edu

News about new gopher servers and software, subcribe to the gopher-news mailing list:
gopher-news-request@boombox.micro.umn.edu

The most recent releases of gopher software is available via anonymous ftp from:
boombox.micro.umn.edu in the /pub/gopher directory.

1. Paper presented at IASSIST 94 in San Francisco, May 1994



IASSIST

INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICE AND TECHNOLOGY

• • • •

ASSOCIATION INTERNATIONALE POUR LES SERVICES ET TECHNIQUES D'INFORMATION EN SCIENCES SOCIALES

The International Association for Social Science Information Services and Technology **(IASSIST)** is an international association of individuals who are engaged in the acquistion, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information system specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompases hard copy as well as machine readable data.

Paid-up members enjoy voting rights and receive the **IASSIST QUARTERLY**. They also benefit from reduced fees for attendance at regional and international conferences sponsored by **IASSIST**.

Membership fees are:

Regular Membership. $40.00 per calendar year.

Student Membership: $20.00 per calendar year.

Institutional subcriptions to the quarterly are available, but do not confer voting rights or other membership benefits.

Institutional Subcription:

$70.00 per calendar year (includes one volume of the Quarterly)

Membership form

I would like to become a member of IASSIST. Please see my choice below:

- ☐ $40 Regular Membership
- ☐ $20 Student Membership
- ☐ $70 Institutional Membership

My primary Interests are:

- ☐ Archive Services/Administration
- ☐ Data Processing
- ☐ Data Management
- ☐ Research Applications
- ☐ Other (specify)_____

Please make checks payable to IASSIST and Mail to :
Mr. Marty Pawlocki
Treasurer, IASSIST
% 303 GSLIS Building,
Social Science Data Archives, University of California, 405 Hilgard Avenue, Los Angeles, CA 90024-1484

Name / title

Institutional Affiliation

Mailing Address

City

Country / zip/ postal code / phone