

-----  
Jacobs, Jim. "Providing Data Services for Machine-Readable Information in an Academic Library: Some Levels of Service." The Public-Access Computer Systems Review 2, no. 1 (1991): 144-160.  
-----

## 1.0 Introduction

Many libraries are facing two trends that are moving them closer to providing services for electronic information products: (1) information in electronic formats is becoming more plentiful, diverse, and obtainable; and (2) a growing number of library users want--and demand--access to information in electronic formats.

One need not look far to find examples of these trends. The proliferation of CD-ROM's in the U.S. government depository program is a good example [1]. Other examples include the availability of electronic journals [2] on floppy disk and through electronic mail delivery, commercially available databases of images and maps, and the wide variety of numeric data files available on computer tape from all levels of government, private vendors, and data archives such as the Inter-University Consortium for Political and Social Research (ICPSR) [3].

Similarly, faculty and students are quite likely today to have easy access to personal computers or powerful workstation-class machines and to feel more comfortable with information in electronic form. As a result, many library users already prefer to have the information they require in a machine-readable format, rather than in paper form.

How can libraries deal with these products and the demands for service that go with them? To address this, I will list some examples of different kinds of services that a user of electronic information might consider important and that a library might consider offering [4].

The four lists are "Levels of General Data Services," "Levels of Computing Services," "Levels of Library Data Services," "Levels of Reference Data Services." (For the purposes of this paper, I have defined "data service" as any kind of service for electronic information.) The lists focus on academic support services, especially library services, for nonbibliographic electronic information products, which I will refer to as "machine-readable information" [5] or "data files."

I specifically exclude bibliographic information products in this

discussion for two reasons. First, there is ample literature on dealing with electronic bibliographic information in libraries. Second, although many libraries now have experience dealing with bibliographic files (e.g., online public catalogs, bibliographic CD-ROM databases, and online bibliographic vendors such as Dialog and BRS), nonbibliographic data products provide different challenges.

Examples of the kinds of products which fall into the nonbibliographic category of machine-readable information include the following: (1) numeric data such as census information, results of survey research, and economic time-series; (2) cartographic data such as census TIGER files; (3) image data such as photographs and satellite images; (4) and textual data such as the full texts of literary works.

Three assumptions or themes underlie these lists.

First, most libraries can provide some kind of service for electronic information without attempting to provide complete service for all conceivable combinations of users and electronic products.

Second, libraries should not avoid dealing with these data resources because of their formats.

Third, it is not necessary, and probably not desirable, for a library to attempt to provide "full service" for machine-readable information on its own. Different campus units (e.g., computer center, survey research centers, and academic departments) might each provide some services which complement those provided by the library and each other. Together, these units may be able to provide better service than any single unit could individually. It is important to analyze one's local academic and computing environment in order to best fit the services into that context [6].

+ Page 146 +

There is one important caveat before we begin looking at levels of service. Although the discussion is intended to be "generic," it is important to remember that every situation is different. The examples used here are just that--examples. They are not intended to be prescriptions for service or suggestions for strategies to follow. Rather, they are reference points that may be used to reflect on one's own situation.

## 2.0 Levels of General Data Service

What kinds of services might a user of machine-readable information expect? What kinds of services might an organization (not necessarily a library) provide? The following is a list of six levels of service which attempt to answer these questions.

Although these levels are not intended to be prescriptive, levels one through three are basic services that should be provided if a campus is to have any level of data services at all. The primary purpose of this list is to help identify what services are already being provided and to help select those services that the library might provide. The levels are listed somewhat hierarchically. Higher levels tend to be more complex or require more staff; they often build on services provided at lower levels.

### 2.1 Level One: Pre-Acquisition Services

There are several things which need to be done before any machine-readable information is acquired. These include (1) receiving requests for data; (2) helping users identify which data files are required; and (3) identifying different sources, formats, and costs of data. On campuses where survey data are important, there should be an ICPSR membership. That membership requires annual funding and a person to serve as the ICPSR Official Representative to process all requests for data and handle communications with the Consortium. That person should also promote ICPSR membership and services on campus.

+ Page 147 +

### 2.2 Level Two: Data Acquisition Services

Once a campus has made a decision to acquire data files someone must insure that they are compatible with local hardware and software and that orders are placed accurately. Next, tapes and other media as well as codebooks or documentation for the machine-readable information must be received and processed. Records of orders placed and received must be kept and bills must be paid. Finally, someone must notify the requester that the data files have been received and provide for physical access to computer tapes (or other medium), codebooks, and technical specifications needed by users to access the data files on their storage medium.

### 2.3 Level Three: Data Access Services

It is important that all authorized data users can easily learn what data files are locally available and how to gain access to those files. Some kind of list or catalog of available data files must be provided, along with the codebooks and technical documentation that will allow authorized users to gain access to these files.

### 2.4 Level Four: Basic Data Advisory Services

Once data files have been acquired, are users "on their own" or will there be consulting services available? Data consultation

or advisory services would require staff who are familiar with the contents and structure of studies, can refer users to particular studies, help users interpret codebooks, and explain how data files are laid out on the storage medium.

## 2.5 Level Five: Data Analysis Advisory Services

Another level of advisory or consultation service would involve staff familiar with statistics, statistical software, and particular academic disciplines. These staff members could advise users on appropriate statistical procedures, help users choose appropriate statistical software, write statistical programs, interpret results, debug statistical programs, and debug analytical procedures.

+ Page 148 +

## 2.6 Level Six: Comprehensive Data Analysis Services

This level of service would do everything for the user. Staff would analyze data as requested and deliver finished output to users requesting analytical products such as charts, graphs, measures of significance, and cross-tabulations of variables. It should be noted that, although this is a rare service to find in an academic setting, it is not unknown. As is true of all these levels of service, some users will expect such service. It is only prudent to anticipate user requests for service and have a clear policy delineating those services that can and cannot be provided.

## 3.0 Levels of Computing Services

Providing services for machine-readable information does not have to include providing computing services, but users of machine-readable information must use computers and providers of data services should be aware of whom is providing computing services for those users.

In these days of smaller, faster, less expensive computers, it is increasingly common for individual users to have adequate computing power on his or her desktop. Libraries interested in equal access to information should question whether the fact that many people have their own computers changes the library's commitment to those who do not. Even if users have their own computers, they must somehow obtain data files they want in formats compatible with their machines and then manage to load them physically into their machines. These processes require some level of computing service on the campus.

In general, there are four basic computing resources, in addition to human resources, that must be provided in each of the levels listed below. These resources are hardware, software, computer "cycles" (i.e., the computer actually performing a task), and

delivery or storage medium. The primary issue is who will provide these services.

+ Page 149 +

### 3.1 Level One: Data Storage Services

Obviously, data files must be stored somewhere on some medium. Will the data files be online or will they require loading or mounting? If stored "offline," who provides the service of loading the files into a host computer? Is a proper storage environment for the chosen medium provided? Will backup copies of files be made? Will someone check whether the files were received in the format ordered and that they were received accurately without errors?

### 3.2 Level Two: Copying and Subsetting Services

Will users be able to copy data files, either whole files or parts of files, to their own account, machine, or disks/tapes? Who will provide instructions for how to do this? Who will provide the hardware, software, and computer cycles for this work? Will the user move data across a network or within a single machine?

This kind of service fits nicely into a traditional library model. For instance, much like a user checks out a book or copies an article from a journal, he or she might copy data of interest from a CD-ROM onto a floppy disk and take it home. With the drop in prices of CD-ROM drives, some users may have their own drives and libraries may want to consider checking out CD-ROM disks. What the user does with the data and what computer resources he or she uses might be of no more concern to the library than whether a user reads a book under a tree on the campus commons or at a desk in a dorm room.

### 3.3 Level Three: Data Retrieval Services

Some data files will be used simply to retrieve a quick fact, a table, a single image, or a brief excerpt of text. Again, who will provide the instruction and computer resources for this service? Many libraries may find it convenient and possible to provide this sort of service for some products distributed on CD-ROM, and it is certainly manageable because no single patron uses any one machine for very long.

+ Page 150 +

Unfortunately, there are many other complicating issues. Here are some example issues. Is there common software available for all files that are needed to answer users' questions or must library staff learn multiple software products in order to help patrons? Are data files documented sufficiently so that users

understand the meaning of the answer they have retrieved? Are librarians sufficiently familiar with the data files on hand to refer library users to the right files in an accurate and efficient manner?

### 3.4 Level Four: Data Analysis Services

One important reason for distributing information in machine-readable format is so that the raw data can be analyzed. Whether this involves performing statistical analysis on a complex census file, overlaying cartographic data over satellite images, or finding word frequencies in a text file, users want data in machine-readable form so that they can manipulate them. A simple analysis might be no more than creating and sorting a list of counties with high per capita income. A more complex analysis might involve performing an advanced statistical procedure on a data file of many variables and thousands of observations. Even simple analysis may take quite some time on a personal computer. Advanced analysis may simply be inappropriate on any but large mini- or mainframe computers. Any kind of analysis will require appropriate software (e.g., statistical, textual, or geographic software).

While libraries might want to provide computers for some kinds of analysis, they may have to develop policies defining appropriate use. Libraries considering providing analytical computing services should realize that they are, in effect, considering becoming a computing center.

## 4.0 Levels of Library Data Services

Let's explore the kinds of data services that a library might provide. This list may serve as one possible application of the more general levels listed above. The numbering of these levels does not necessarily correspond to the "Levels of General Data Service" listing. This section serves as an example of how different organizational situations are different and require their own solutions.

+ Page 151 +

### 4.1 Level One: Passive Referral Services

The lowest possible level of service (other than no service at all) may already be in place; however, minor service enhancements may be desirable.

This level of service can be provided by a knowledgeable reference staff equipped with an adequate supply of printed materials and access to appropriate online databases that list sources of and collections of machine-readable information. No additional staff or separate service center is required, although some staff training and some additional reference tools may be

needed.

At this level of service, staff determine whether a certain kind of information exists in machine-readable form, where it can be obtained, and how much it costs. This level of service is passive in that it does not actively seek patrons or users of machine-readable information, but simply responds to questions by referring patrons to vendors or other collections.

Normal online bibliographic searches for library users could be broadened to include databases that list machine-readable information (e.g., ERIC, RLIN, NTIS, and ICPSR Guide). The service could be further enhanced by adding "codebooks" (i.e., descriptions of the contents of machine-readable data files) to the general collection of books. These codebooks would aid researchers in identifying useful data files, and they would often provide actual useful data themselves, such as frequencies of response to individual questions.

#### 4.2 Level Two: Active Education and Referral Services

This level adds the education of users and promotion of services to level one activities. It is the active counterpart to level one. Level two aspires to make users, and potential users, of the library aware both of the existence of information in machine-readable form and of the services which the library can provide in identifying such products. Education and promotion may be in the form of user instruction classes and seminars, special workshops on "new" sources of information for particular subjects, informal contacts between librarians and faculty, and newsletters.

+ Page 152 +

#### 4.3 Level Three: Data Cataloging Services

On campuses where there are already machine-readable information collections outside the library, level three is a very important and fairly easy step to take to improve data services. There may be a collection of data files on campus in a computer center, data archive, social science research center, or even a department or faculty office, but these data files are not accessible to all potential users because there is no central listing of them. The library might offer to catalog the data files or the codebooks, or both, and add those cataloging records to its online or card catalog. If there is no current easy access to the code book collection, the library might also offer to house and maintain it, or the library might choose to buy copies for its collection, adding yet another access point.

#### 4.4 Level Four: Data Acquisition Services

This level involves the addition of machine-readable information

to the library collections. Although this may not seem to be an immediate prospect, libraries that are government depositories are already having to decide whether or not to accept machine-readable depository items.

Decisions will have to be made as to: (1) which media will be collected (e.g., tape, floppy disks, compact disks, and video disks); (2) which formats will be collected (e.g., for tapes: what densities, number of tracks, and character modes); (3) what level and kind of cataloging or other bibliographic access will be provided; (4) where machine-readable information will be stored; (5) what criteria will be used to select and acquire machine-readable information; (6) what kind of access will be available; and (7) who will have access. It would be wise to write a formal collection development policy statement for machine-readable information in order to both address these issues within the library and to communicate to faculty and other users how much, or how little, the library can do.

+ Page 153 +

#### 4.5 Level Five: Data Consultation Services

Deciding to acquire data files raises the question of what other kinds of services will be provided for the information acquired. Data consultation services can be offered when there is machine-readable information on campus, whether it is acquired by the library or by another agency on campus. These services also can be offered on widely differing levels.

#### 4.6 Level Six: Archival Services

It is fairly easy to buy a few data files on tape and, in one way or another, add them to the library collections. However, it will take much more effort and time to archive campus machine-readable information. This level of service might well be omitted as it is not a necessary prerequisite to higher levels of service.

This level could be seen as an extension of level three or level four services. As an extension of level three services, archiving data files would mean assuming responsibility for a large collection of data files all at once. This is unlikely to happen unless there is currently no formal location for the collection. For example, it is only informally housed in a faculty office. As an extension of level four services, archiving data files would involve storing and making accessible files acquired or produced by individuals on campus.

Just as in a traditional archive, it is very important to have a clear statement of what you will accept and what you will not. Documentation, or the lack of it, can be a particularly sensitive issue when evaluating locally produced data files.

#### 4.7 Level Seven: Data Analysis Services

Not every library will want to provide all levels of service, but data analysis services may be the least likely to be offered. In general, libraries do not offer this kind of service even for printed materials with the exception of "ready reference" questions. The more complex the question, the less likely most libraries are to provide answers. The obvious example of this service orientation in terms of printed materials are medical and legal questions, which for ethical and legal reasons are virtually never answered. Another example is that few libraries answer questions which involve interpretation of tables of statistics.

To continue the analogy with printed sources, data service at level five would be comparable to helping someone locate the appropriate volume of the printed census, the appropriate table in that volume, and the appropriate explanations of how the data were collected and how they are presented, but would leave the user to read, interpret, and choose which numbers actually answer his or her questions. By contrast, level seven service would take a question or a precise request for data analysis from a user and provide that user with an answer to the question or a customized product of data analysis. This would require all the sophistication of the other levels, plus a more experienced staff and more staff time than any other level.

#### 5.0 Levels of Reference Data Services

Assuming you have some data files and you want to provide some sort of reference service for those files, where do you start? Or, more appropriately, where do you stop? Here are some examples of levels of reference service. Once again, these examples, which are based on an academic library context, are meant to help guide your thought and help you plan within your own context. They are not meant to be definitive guidelines.

#### 5.1 Level One: Data File Identification Services

If someone asks "Do you have the PSID?" [7], your reference desk staff should be able to understand the question, find out if the PSID is on campus, and, if it is, where it is, who has access to it, and what the access procedures are. Also, in a case like this one, you'd want to be able to identify which parts of PSID you have and if they are in some special format or not.

Adding catalog records for your data file holdings to your online or card catalog can accomplish most of this. However, general awareness of data files is also necessary. Special guides that list your local holdings and guides explaining how to access data

files would also help. Such guides might be created by the library, the computer center, a research center on campus, or a combination of such organizations.

## 5.2 Level Two: Basic Data File Recommendation Services

If a patron asks "Do you have some statistics on income?", you could, if you had cataloged machine-readable information, search your catalog by subject and find National Longitudinal Survey, Current Population Survey, the Census of Population and Housing, the Survey of Income and Program Participation, and numerous other entries. But how helpful is that?

A more useful service might provide at least some guidance on the differences among these files. Reference staff should know the difference between aggregate data and micro-data [8], and they should comprehend to difference between cross-sectional and longitudinal studies [9]. They should understand what a panel study [10] is and be comfortable talking with users about sample size, choice of sample, level of observation (e.g., household and individual), geographic detail, and so forth [11].

+ Page 156 +

## 5.3 Level Three: Advanced Data File Recommendation Services

Providing service for large data files is somewhat like providing service for a collection of manuscripts. Typically, a data file will record information on dozens, or even hundreds, of topics; and, typically, few of these topics are indexed in traditional library catalogs. Just as an archivist may remember a letter from an ex-slave buried in the papers of an Ohio school teacher, a data archivist may remember that a particular poll asked a question about day care availability for single parents. Such familiarity, which comes from reading codebooks as data files are acquired and from working closely with data users as they use data files, increases the access points to a collection.

## 5.4 Level Four: Data File Use Advisory Services

Data archivists learn about limitations of data files and hear about their problems by working with users and data and by talking with other librarians. An example of a data file limitation is data that are stored in a special format and require a specific piece of software for access. Researchers who use a data file will be well aware of some problems associated with it, but other problems will not be so well known. As librarians acquire this kind of knowledge, they can help users by sharing it with them.

For example, several problems with the content of U.S. foreign trade data were discussed at a recent meeting of the Association of Public Data Users: (1) for several years, foreign trade data

had a "carryover" problem (data reported for one month actually included trade from earlier months); (2) the change to the Harmonized system of classifying industries makes it very hard to compare current data with older data; (3) exports are not counted as carefully as imports; (4) aggregate figures are revised, but industry level figures are not [12]. It is apparent how this information, not all of which is documented, would be very helpful to a user of U.S. foreign trade data.

+ Page 157 +

## 5.5 Level Five: Data Extraction Services

When librarians help novice or occasional data file users to obtain subsets from large data files, their familiarity with how particular data files are organized and arranged is important. Even if computer or programming assistance is not provided, it is helpful to understand the different data structures so that you can help the user identify if the needed data need is available and if they will be easy to extract. For example, each record in the Citibase database is a time series; therefore, it is very easy to extract time series from Citibase [13]. In other data files, time series data may be available, but it may be embedded in variables, with each record consisting of an observation for a person or a household. Extracting a time series from such a data file would be a much more difficult process.

## 6.0 Conclusion

Nonbibliographic machine-readable data files provide many challenges for libraries and their campuses. Many of these challenges can be met by combining the resources and skills of different units on campus in order to provide a coherent service. The key to providing such a service is analyzing local resources and needs and making wise choices among a wide range of possible services. Libraries have an important role to play in the provision of data services, and they can provide many data services with little or no change to staffing or other resource allocations.

Librarians interested in more information about machine-readable information and data services should investigate membership in the ICPSR, IASSIST [14], and APDU.

+ Page 158 +

## Notes

1. Depository Library Council, Subcommittee on Electronic Distribution, "Preliminary Report, March 9, 1988," Administrative Notes 9, no. 8 (May 1988): 20-26; and U.S. Government Printing Office, GPO Special Survey 89-300 (Washington, D.C., U.S. Government Printing Office, October 1989).

2. Ron Eisner, "Publishers Work Toward Starting Reputable Online Science Journals," *The Scientist* 5 (4 March 1991): 4-5.

3. For a list of studies available from ICPSR, see the annual publication: *Guide to Resources and Services*.

4. Earlier examples of lists of levels of service appear in the following sources: Howard D. White, "Libraries and Access to Social Science Data," in *Reader in Machine-Readable Social Data*, ed. Howard D. White (Englewood, Co. Information Handling Services, 1977), 175-194; Laine G. M. Ruus, "The University of British Columbia Data Library: An Overview," *Library Trends* 30 (Winter 1982): 397-406; Edward P. Bartkus, "Use of Numeric Databases in Reference and Information Services," *Drexel Library Quarterly* 18 (Summer-Fall 1982): 205-219; JoAnn Dionne, "Why Librarians Need to Know About Numeric Databases," in *Numeric Databases*, ed. Ching-chih Chen and Peter Hernon (Norwood, NJ: Ablex Publishing Corporation, 1984), 237-246; and Ann S. Gray and Sue A. Dodd, "The Roles of Libraries and Information Centers in Providing Access to Numeric Databases," in *Numeric Databases*, ed. Ching-chih Chen and Peter Hernon (Norwood, NJ: Ablex Publishing Corporation, 1984), 247-262. The lists presented here are derived from these earlier works, personal experience, and numerous communications with other librarians attempting to deal with these issues.

5. An excellent overview of nonbibliographic databases is provided in: RASD/MARS Committee on Nonbibliographic Databases and Data Files, "Information Sheet I: What Are Nonbibliographic Databases," *RQ* 26 (Spring 1987): 280-284.

+ Page 159 +

6. Diane Geraci, "Categorizing Your Local Environment" (Paper presented at Management of Machine-Readable Social Science Information Workshop, ICPSR Summer Program in Quantitative Methods of Social Research, Ann Arbor, MI, August 1990).

7. James N. Morgan, "Panel Study of Income Dynamics" (Ann Arbor, MI: Survey Research Center, Inter-University Consortium for Political and Social Research, 1989). (Computer file)

8. Aggregate data are those which have been created by combining values for a number of individual observations into a larger unit. An example would be census data files which contain values that are totals for geographic areas such as blocks and counties, without the values for each individual respondent to the census. Micro-data are those which contain values for individual respondents.

9. Cross sectional data provide observations on a group sample at a particular point in time. Longitudinal studies are studies across time.

10. A panel study interviews the same group of individuals (the

"panel") several times over a period of months or years.

11. A good source of definitions of survey research concepts is P. McC. Miller and M. J. Wilson, *A Dictionary of Social Science Methods* (Chichester, England: John Wiley and Sons, 1983).

12. The Association of Public Data Users (APDU) is an organization of users, producers, and distributors of federal, state, and local government statistical data. The executive director is Susan Anderson, Princeton University Computing Center, 87 Prospect Ave., Princeton, NY 08544.

13. Citibase: Citibank Economic Database (New York: Citibank). (Computer file)

14. IASSIST: The International Association of Social Science Information Service and Technology.

+ Page 160 +

About the Author

Jim Jacobs  
Data Services Librarian  
University of California, San Diego  
Central University Library, 0175-R  
9500 Gilman Drive  
La Jolla, CA 92093-0175  
JAJACOBS@UCSD.EDU

-----  
The Public-Access Computer Systems Review is an electronic journal. It is sent free of charge to participants of the Public-Access Computer Systems Forum (PACS-L), a computer conference on BITNET. To join PACS-L, send an electronic mail message to [LISTSERV@UHUPVM1](mailto:LISTSERV@UHUPVM1) that says: SUBSCRIBE PACS-L First Name Last Name.

This article is Copyright (C) 1991 by Jim Jacobs. All Rights Reserved.

The Public-Access Computer Systems Review is Copyright (C) 1991 by the University Libraries, University of Houston, University Park. All Rights Reserved.

Copying is permitted for noncommercial use by computer conferences, individual scholars, and libraries. Libraries are authorized to add the journal to their collection, in electronic or printed form, at no charge. This message must appear on all copied material. All commercial use requires permission.  
-----