Robbin, A. (1977). The pre-acquisition process: A strategy for locating and acquiring machine-readable data. *Drexel Library Quarterly*, *13*(1), 21-42.

Forthcoming Issues
of the
Drexel Library Quarterly
1977

Drexel Library Quarterly

A publication of the
Graduate School of Library Science
Drexel University

January 1977 Vol. 13 No. 1

Machine-Readable Social Science Data

Howard D. White
Issue Editor

This issue of the *Drexel Library Quarterly* contains reports by representatives of the new breed—persons active in the data archive movement. It is the first issue wholly on their concerns to appear in the library press.

Notes

1 Cf. Elizabeth Herman and John Byrum, eds., *Final Report of the Catalog Code Revision Committee Subcommittee on Rules for Cataloging Machine-Readable Data Files* (Chicago: American Library Association, 1976). ED 119 727. See in particular the types of files mentioned in the appendixes.

2 James Martin, *Computer Data-Base Organization* (Englewood Cliffs, N.J.: Prentice-Hall, 1975), pp. 10-16 *passim.*

3 Ibid., pp. 11, 45.

4 John D. Byrum and Judith S. Rowe, "An Integrated User-Oriented System for the Documentation and Control of Machine-Readable Data Files," *Library Resources & Technical Services* 16 (Summer 1972): 338-346.

5 Lucinda D. Conger, "Data Reference Work with Machine Readable Data Files in the Social Sciences," *Journal of Academic Librarianship* 2 (May 1976): 60-65.

6 Martin, pp. 14, 16, 45.

7 Ibid., pp. 19-29.

8 For an interesting proposal of how random sampling might be used in bibliographic searching, see Marcia Bates, "Rigorous Systematic Bibliography," *RQ* 16 (Fall 1976): 21-22.

9 Institute for Survey Research, Temple University, "Final Report for a Survey of Student Utilization of Paley Library" (August 1974), pp. 16-17. I wish to thank Arthur T. Hamlin, director of libraries at Temple, and Richard B. Vanderveer, who directed the Paley study for the Institute for Survey Research, for use of their materials.

The Pre-Acquisition Process: A Strategy for Locating and Acquiring Machine-Readable Data*

Alice Robbin

Introduction

The special library, in contrast to the research or academic library, has been created to respond most directly and immediately to its special clientele. The special library differs from other libraries not only in the nature of its collection, but in the size and quality of the staff and the nature of the interaction between staff and clientele. This means that the collection reflects as closely as possible the clientele's activities, that the library staff's substantive and technical training provides the clientele with an extensive knowledge of the nature of the library's collection, that the staff has a comprehensive knowledge of resource materials described in its reference collection and potentially available elsewhere, that its philosophy regarding user services includes attention to special requests for materials and information, and that it has an opportunity to maintain flexibility regarding its various policies.

Walford and Batten have suggested that the purpose of special libraries and information services is to "provide each user with the information he needs . . . at the time when he needs it, ensuring that material for which there is a known demand or which is likely to be required at short notice is either in the library collection or readily obtainable from other sources."[1] With regard to acquisition of new materials, special attention to a particular client's needs may mean that the library staff will perform an extensive search through the library's reference collection and will interact with many potential sources of these materials.

The more traditional libraries have a long history and (thus) a formalized communications infrastructure essential to obtaining in-

*The author wishes to thank Kerry Schmitt, program consultant at the Social Science Data and Program Library Service, University of Wisconsin-Madison, for his suggestions for the flowchart describing the pre-acquisition process and Kathleen Heim of the library school, University of Wisconsin-Madison, for her helpful comments.

formation and materials easily and efficiently. But the data dissemination system is relatively immature—the oldest archives are no more than 16 years old and the majority of the data libraries have been established in the last 5 years. The result is that the communications infrastructure of the data dissemination system is nowhere as well developed as that of the more traditional libraries. Peters has noted that the problem regarding systems networks is that of establishing network "memories" or information about data in order to meet special requests for particular data.[2] Due to an immature data dissemination system, there are inadequate network memories for organizing and transmitting information about data, and therefore for making visible and accessible the holdings of data libraries, archives, and other organizations. Thus, access to information on data resources becomes a critical problem for members of the system when they attempt to provide information to their clientele in an efficient and timely fashion. Because information about data resources is not easily organized to facilitate their visibility, the search activities undertaken on behalf of a client to locate data and determine their availability are in general rather protracted and inefficient.

The intent of this article is to describe how the social science data library responds to requests for particular data when the data are not cataloged in the data library's collection. The author defines those activities undertaken to locate a potential resource as *the pre-acquisition process.* The pre-acquisition process begins when the library staff and client have been unable to locate relevant data in the library. Depending upon the nature of the request, time constraints imposed on the client (to produce an analysis based upon these data), and status of the client in the university setting, both staff and client undertake a search to locate the data the client needs. Together they examine the library's reference collection and literature in the field, seek information from experts in the field, and correspond with libraries and institutions which might potentially hold the data or be able to provide assistance in locating the data.

The decision to undertake these activities during the pre-acquisition process depends upon a number of factors. They include the library's mandate regarding the nature of the collection and of its clientele, the level of flexibility in its policies, the quality of the staff, the degree to which the staff is committed to providing user services, the quality of its reference collection and ser-

vices, and constraints on library resources. Although constraints on the library's resources may indeed be a major factor in determining whether to undertake a search and at what level and to what extent the services shall be provided, the first part of this article presents a brief description of only five factors: the nature of the collection, the impact of an acquisitions policy upon the pre-acquisition process, staff, the interaction between staff and client, and reference services.* The second part describes the pre-acquisition process carried out at the Social Science Data and Program Library Service (DPLS) at the University of Wisconsin-Madison as an illustration of locating data to meet a special client request. The author believes that a description of these activities should have utility not only for data libraries, but for other libraries for formalizing the process of meeting special requests.

Part One: Five Factors Which Influence the Pre-Acquisition Process

Nature of the Collection and Clientele

The special library's collection consists of information acquired for use by or on behalf of the library's clientele. The social science data library has a collection of information in machine-readable form which is *used* (not necessarily generated) by social scientists.[3] These materials include census information, sample surveys, opinion and market research polls, bibliographic and textual data bases. Information used by social scientists broadens the nature of the collection to include not only the traditional subject areas of sociology, political science, and economics, but history, education, psychology, anthropology, and geography, in fact any materials used by those who call themselves "social scientists."

---

*There is no doubt that constraints of budget, personnel and other resources are critical to a decision to service a library's clientele; but I would suggest that these factors influence the *quality* of the activities more than *what* activities are performed.

The subject coverage of the collection is intensive and/or extensive (comprehensive) enough to meet current and anticipated requirements of the clientele.*

History would indicate that the establishment of a data library was proposed by the *research* community in an academic setting where it was necessary to support a high level of doctoral and post-doctoral research by acquiring and maintaining very specific types of social science machine-readable data. While this would suggest that the social science data library's collecting activity would be at the "intensive level," developments in the 1970s seem to indicate that data libraries have adopted additional policies in response to changes in the academic environment.

In the middle 1960s, few social scientists—researchers and only advanced dissertation-level graduate students—used machine-readable data and the computer as part of their ongoing research efforts, and therefore the data library was a facility made available for a small elite of quantitative and behavioral scientists. Utilization of data and the computer was in the hands of researchers and was directly tied to the number of research dollars which university social scientists received from the federal government. But, by the 1970s, utilization of data and the computer for *classroom instruction* began to be considered a necessary and integral

*Grieder of Stanford University in *Book Selection Policies* suggests the following levels of collecting activity: *intensive level,* whereby the policy is intended to support doctoral and post-doctoral research with a high degree of adequacy and minimum reliance on outside resources, allowing "for an indefinite expansion of the research program, both at the PhD. level and in post-doctoral and faculty research;" *comprehensive level,* whereby doctoral work is expected to be supported with "reasonable" adequacy, but where a large graduate enrollment with many fields of concentration for the doctoral degree will not be adequately supported and where there is more reliance on loans and purchases from other libraries; *beginning research level,* whereby fundamental works of scholarship, everything for undergraduate teaching and a majority of works required for teaching up to the master's level are acquired, and where beyond this level, the response is to individual faculty and student demands; *teaching level,* whereby the collection is maintained to support effective undergraduate instruction; and, *reference level,* whereby materials are selected primarily "because of the relationship of a subject not formally offered to existing instructional or research programs or to serve occasional general needs."[4]

component of the social science learning process, and departments began to offer training in methodology and data handling. The data library began to service increasing numbers of undergraduate and graduate students using the library for coursework purposes. A number of newer data libraries responded to the institutional setting by developing acquisitions policies which reflected the growing emphasis on undergraduate and master's level instruction and by responding to specific demands made by faculty and students ("beginning research level"). In other words, a clientele composed primarily of the research community was enlarged to include a community of undergraduate and graduate students who used the data library primarily for coursework. (And an acquisitions policy had to be developed which would encompass the needs of both the researcher and student.)

For the pre-acquisition process, the nature of the data library staff's response has meant a series of decisions regarding not only how the data would enhance the quality of the library's collection, but also the status of the client and the nature of his request. This is to suggest that not all data needs will be met by the library staff even when an initial determination is made that the data will enhance the collection, but rather that the staff must determine who is requesting the data (e.g., undergraduate, graduate student, or faculty member), and how the data will be used (e.g., research purposes, or limited use and course project).

Acquisitions Policy

Because the special library serves a limited clientele, the library has the opportunity to create flexible policies for acquisition of new materials. Although every library ought to have an articulated policy for acquisitions, the decision whether all the criteria which comprise the policy need be met in every instance of potential acquisition should be based upon consideration of the client's needs and the library's ability to meet these needs. Although it is not the intent of this article to establish the criteria for a data acquisitions policy, it is important to note that in terms of providing good user services, a flexible policy is an essential element in the pre-acquisition process.

At a social science data library, the issue of an acquisitions policy is raised at the outset of the pre-acquisition process, even before the search is undertaken by the library staff and client. This is because oftentimes the data library staff is unable to assure its client that the data will definitely be acquired by the library before the data are located and are determined to be available. Locating data does not necessarily mean that data are available. Because most data are not in the public domain, it is often necessary to carry out negotiations with the producer of the file in order to obtain it (e.g., only selected variables may be available; only selected types of analysis may take place; data are still being analyzed). Availability is constrained by cost of acquisition, quality of the documentation and data, transferability from one site to another, cost of duplication, processing and analysis software at the client's site, and technical (computer and data) support staff. Thus, it is possible that staff and client will have performed an extensive search for data, will have located them, will have learned that they are available, and then will be unable to acquire them because consideration of the cost of acquisition and utilization, and the quality of the data and documentation, make it unfeasible to acquire the file.

Quality of the Staff

The special library is usually staffed with individuals having special and in-depth substantive and technical training in the type of collection maintained.* Walford and Batten have noted that "[a special library] staff must be well acquainted with the relevant sources of documents and information, aware of the interests of the organization they serve and must understand their users' needs."[6]

---

*It has been suggested that adequate substantive and technical training would include a master's level degree in one of the social science disciplines, with emphasis on quantitative research and methodology and "hands-on experience" with data (i.e., some analysis and use of the computer for analysis purposes). As Conger has suggested, while knowledge of computer programming is helpful, it is not really necessary;[5] however, for decisions involved in data acquisition and transfer of data, it is important to understand the constraints of accessing a file created at another site and it is important to be able to explain, in a language which the programmer can understand, how the data are structured.

In the social science data library, the substantive and technical training of the staff means that the staff is equipped to provide information about the availability of data and access to them and can provide technical assistance in utilizing the information. The social science data librarian has substantive background in one of the social science disciplines (the background is usually broad rather than deep), is familiar with current research and aware of the quantitative research activities of specialists in the social sciences. He or she is knowledgeable about data resources and their location, the quality of the data and their potential for future research, the transfer of the materials from one location to another, and the preparation of data for analysis.

The extent of information about data resources and the technical expertise of the staff are critical during the pre-acquisition process. Experience is necessary for the extensive information that a good reference librarian can provide.* And the librarian is called upon to *evaluate* the quality of information which might be received. The data library reference librarian must draw upon knowledge (gained largely through experience in the library) of the substantive aspects of particular data bases, of individuals and organizations engaged in the production of the particular data required, and of the quality of production of these data.

---

*It has been suggested that the data library's reference collection be computerized in order to retrieve selected information in a more efficient fashion than by dependence on the librarian's memory. At present, it is not clear how the materials which are potentially useful could be stored and retrieved. As will be shown in Part Two, information which could prove relevant to a client's needs may be located in the body of certain texts, in footnotes and citations, journal and newsletter articles, and "a librarian's notes." It is obviously not possible or useful or efficient to input all information (or even a large amount) into the computer because 1) most information does not prove to be useful, 2) it is enormously costly (development, maintenance, and retrieval costs) to maintain an information storage and retrieval system for any data base, let alone a library's reference collection, and 3) there is no agreement on the search vocabulary which would satisfy a variety of users' needs and thus facilitate retrieval of selected pieces of information. Thus, the library continues to rely on the random access memory of the experienced librarian, a memory which is in almost every instance far more efficient than a computer's.

Staff-Client Interaction

The special library emphasizes a personal interaction between library staff and client. The client group expects that the library staff has a good understanding of the nature of its needs and that the staff can respond to these needs with a minimum of explanation. Because a special library is designed to permit more intensive and extensive library staff-client interaction, there is the expectation on the part of both the staff and client that time will be spent understanding the nature of the client's problem and attempting to meet his needs to the fullest extent possible.

In terms of the pre-acquisition process, more intensive library staff-client interaction also permits "a more direct participation by the users in the [pre-] selection process [which] may be the only way to make the fullest use of available sources."[7] The search process undertaken by the data library staff includes full participation by the client. The client's search sometimes reveals formerly unknown resources. In fact, successful locating of data resources is often the result of the client's own search.

Reference Collection and Other Resources

No library can be expected to collect everything. Indeed, not only do limited storage and working areas make it impossible for a collection to expand indefinitely (even with technological advances in the storing of information in smaller and more compact areas), but it is important to ask whether a library *should* acquire everything. If the policy is not to acquire everything (although anticipating client needs is a principle which should be adhered to if a library is to be dynamic), then it becomes even more essential that the special library contain the reference resources and special information that are necessary for locating a potential acquisition.

For a social science data library, it is more important to create an extensive reference service and for the staff to be well informed about available data resources than to acquire them before there is an expressed need. The reference collection in the social science data library should contain the listings and documentation

on the holdings of other data libraries or archives, newsletters* and journals** which describe completed or in-progress research regarding social science-related data collections, bibliographies, abstracts and indexes which describe organizations carrying out social science research based on analysis of machine-readable data or carrying out data collection as a function of their mandate. The data library staff itself may compile inventories, indexes, abstracts or digests of pertinent literature—in particular, monographs, journal articles, and papers presented at meetings and conferences which describe the results of research based on analysis of machine-readable data. The staff will collect, on a regular basis, codebooks and study descriptions of available data files and, on an irregular basis, methodology textbooks, book reviews, newspaper clippings, publishers' announcements, and fugitive papers in which machine-readable data have been cited.

The data library may request that librarians located in other special or academic libraries at their particular institution perform selective dissemination of information related to the data library's special needs. (This is a cooperative project, maintained on an informal basis, which has the potential of reaping benefits for all librarians concerned, especially because it extends the communications network essential for good reference services.) The selected information, while not necessarily cataloged in the data library's reference collection, becomes part of the materials to which the data librarian refers when special requests are made.

The social science data library staff faces an enormous task if it is to respond adequately to special requests for materials or information. As Nasatir has noted:

---

*Many newsletters are published on an irregular basis and circulated among a few individuals. It is important for the reference librarian to be aware of what is produced in this form and to include as many of them as possible; it is often within them that information is located. Some of these include *SS Data, Historical Methods Newsletter,* and *Federal Statistics Users Conference Newsletter.*

**A journal collection at a social science data library includes those journals which describe research related to quantitative social science research (e.g., *Journal of Interdisciplinary History, Computers and the Humanities, Social Science Information*).

> Every hour of every day the production of data from academic, commercial and governmental sources continues at an increasing pace. . . . In addition to traditional sources of statistics such as government and industry, the past four decades have also produced a continuously growing amount of data on an extraordinary variety of topics due to steadily growing interest in public opinion polling and market research. . . . In the post-World War II era, an increasing amount of money and equipment [has become] available for social science research and for social scientists, not only for the scientific and commercial interests, but also in response to an ever more pressing quest for data to satisfy the needs of government and the requirements of rational planning.[8]

What this has meant for the data library is that most of the resources which could potentially be utilized by its clientele have neither been archived nor referenced nor placed in some way in the public domain. To respond adequately to a client's particular needs, therefore, requires far more than the usual amount of search capability. The librarian cannot rely only on the library's reference collection, but must rely on the client to conduct a parallel search, on other experts in the field, and perhaps most importantly, on the library staff's own time-consuming search (and imagination, it should be added!), which might take weeks to months and lead the librarian to unexpected realms of federal agencies and bureaus and the globe. The extensiveness of the search for materials for a client may be limited only by the library's resources and time constraints imposed by the client's own needs.

Part Two: A Search Strategy for Locating Data

Introduction

When a client arrives at the data library, the first step is to ascertain whether the data he needs are in the collection. The client adopts one of two search strategies. He may decide to conduct his own search of the library's collection, continuing until the search strategy proves ineffective. Or the client directly queries the library staff about which data files contain the variables needed for the intended project. Most clients choose the second strategy due to 1) their level of confidence in their ability to locate information which appears to be organized in a nontraditional way, 2) their lack of expertise in the area of data handling, 3) their belief that the library staff will understand their needs, and 4) their view that the staff is there to facilitate the search process.

The query negotiation process, which takes place when client and staff interact, reveals the client's data needs. The degree of information provided by the client depends on a number of factors: 1) the client's knowledge of his needs; 2) his level of sophistication with data and analysis via the computer; 3) the previous information that the client has about the library staff and its capacity to understand his query; and, 4) his previous use of the data library's facilities. During the query negotiation process the staff will learn about the client's research design and variables needed for analysis purposes, for what reasons the research is being conducted, and for what purposes (i.e., coursework or "long-term" research).

With the knowledge of the needs of the client and his "time-frame," the library staff begins a search of the collection. The data librarian recommends examining data files which may contain the necessary variables for analysis. The data librarian's search through the holdings is usually at the variable level because the client is concerned with locating multiple data items within a particular study. (All the data items must be contained within a particular study in order that the study be useful to the client.) Should the search by the library staff and client not reveal the needed variables, it is at this stage that the pre-acquisition process begins.

Pre-acquisition Process: Overview

There is a high probability that the staff will initiate a search to locate data if:

1 The project involves research (or has more long-term implications) or the client has invested a great deal of energy in defining the project needs and could not easily redefine the research design; and,

2 There are no or few time constraints on the client for locating the data; and/or,

3 The client is an advanced graduate student or faculty member or researcher.

There is a high probability that the staff will recommend a redefinition of the study design if:

1 The data needs are ill defined or the study design can be redefined easily or "hands-on experience" is more important than the outcome of data analysis; and/or,

2 There are time constraints for meeting project deadlines; and/or,

3 The client is an undergraduate or graduate student.

A search of the library's reference collection involves an examination of codebooks, study descriptions, catalogs and listings of archival holdings, newsletters and journals, inventories, directories, articles and miscellaneous information which the librarian has collected. The search may be as time-consuming as staff resources permit and as long as it takes for a "hit" to occur (to employ a useful computer term!). Obviously, limits will be imposed by necessity and library resources are finite.

If the search of in-house reference materials is unsuccessful, the library staff contacts the client and requests that he perform a search through other reference sources. If the search by the staff and client is successful, the next step is to determine data availability. Many data files, although described in source materials, may not be available to either the library or client.

The library staff determines whether it is more appropriate for it or the client to write a letter asking whether the data are available. Political considerations manifest themselves at this stage: Many generators of data are reluctant to release their data to a library or archive because they believe (incorrectly) that archiving the file means to place it in the public domain without restrictions, or because they have not completed analysis of the file, or, sometimes, because they believe that archiving the file will make possible the analysis of their data and thus subject their findings to

question. But some will release their file(s) to particular individuals on a restricted basis if the client's research design and use of variables are fully described.

If the in-house library and follow-up client searches do not reveal any data sources of potential utility, the library staff and client begin a major effort to locate the data. This means telephone calls and letters to archives, to individuals who are specialists in the area of the client's research, and to specialists within archives. The search is as long as the resources of the library staff and the client permit.

Should the search be unsuccessful, the client must redefine his project. If the search is successful and data are available, the library staff must determine whether to acquire the data. If the data are not acquired by the library, the client has the choice of acquiring the data himself or redefining his research project. If the data are acquired by the library, the staff makes a formal request for them.

See Appendix for a detailed description and flowchart of the pre-acquisition process.

Summary

This article has described the activities performed by the library staff during the process preceding the actual acquisition of data on behalf of the data library's client.

The quality of the library staff's performance is influenced by the policies which have been established for creating and augmenting the collection, the background and experience that the library staff brings to its work, the nature of the interaction between staff and client, the extent and quality of responsiveness to a client's needs for materials and information, and the extensiveness of the reference/resource materials available to the library staff within the library and through a network of informal communications of the staff with colleagues in the data archive profession and specialists in the academic social science community.

This article has not discussed other critical issues related to the pre-acquisition process, such as constraints on the library resources, degree of success in locating data, and quality of data and criteria for their acquisition. It is obvious that decisions made during the search will be based not only on consideration of user services, but will also be influenced by issues which were only alluded to in this article. Indeed, the library may not have the resources to do much more than initiate a search by directing a client to potential sources of information, forcing the client to fend for himself. Second, there is a low level of success in locating data. Responses are infrequently made, follow-up letters are necessary, but library resources limit the level of effort which the staff can make. Too often, the client is "forced" to change his study design or collect his own data. Further, the quality of data is a major deterrent to library acquisition. And library budgets are so limited and data so expensive that libraries can make few purchases.

But these are problems better left to future discussions when a more detailed analysis can be made. It is hoped, however, that the preceding discussion has provided some insights into the extensive activities that a data library staff performs when responding to special requests for data sources which are not cataloged in the library's collection.

Notes

1 A. J. Walford and W. E. Batten, "Selection and Acquisition of Library Materials," in *Handbook of Special Librarianship and Information Work*, 4th ed., ed. W. E. Batten (London: Aslib, 1975), p. 72.

2 Paul Peters, "Describing a Social Science Data Information System, Networks and Components," in *The Data Library: Systematic, Structural and Process Problems of Data Access*, ed. Alice Robbin. *SIGSOC Bulletin* 6 (numbers 2 & 3): 11.

3 David Nasatir to IASSIST Data Archive Registry Group members in a memo dated 2 June, 1976.

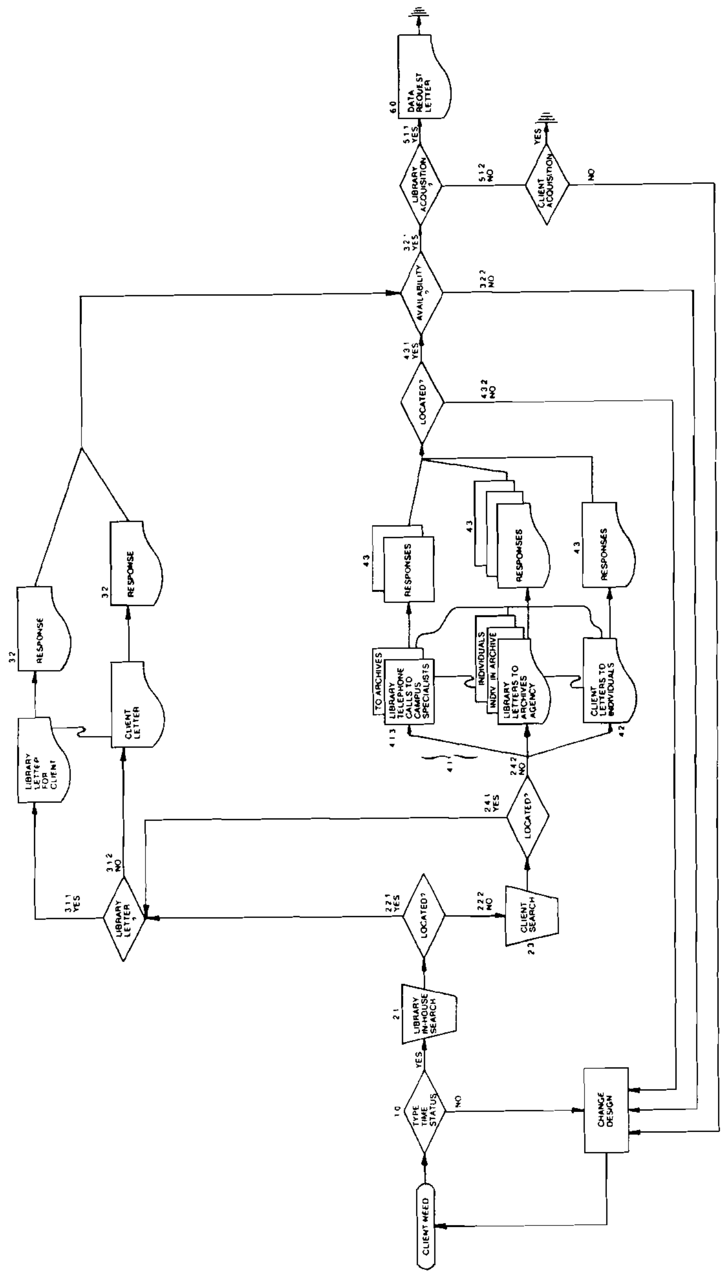
4 E. M. Grieder, ed., "Stanford University Libraries Book Selection Policies," mimeographed (Stanford: Stanford Libraries, December 1970).

5 Lucinda D. Conger, "Data Reference Work with Machine Readable Data Files in the Social Sciences," *The Journal of Academic Librarianship* 2 (number 2): 64.

6 Walford and Batten, p. 72.

7 Ibid., p. 77.

8 David Nasatir, "Social Science Data Archives: An International Perspective," pp. 3-4. Paper presented at the 1973 Congress of the World Association for Public Opinion Research (Budapest, September 1973).



The Pre-Acquisition Process: A Search Strategy for Locating (and Acquiring) Machine-Readable Data

Appendix: The Pre-Acquisition Process in Detail

The following is a step-by-step description of the activities that DPLS performs in the pre-acquisition process and the decisions made along the way which influence these activities. A flowchart describes graphically the process from the point at which the client describes his research design through the actual decision to request the data file. The text describes only the pre-acquisition process; the decisions regarding acquisition are not treated in this discussion.

1.0 **Decision to begin the process.**
Three variables determine whether the library staff will begin a search for data. In order of priority, they are: 1) type of project, coursework or research; 2) length of time that client has to complete project; 3) status of client, undergraduate, graduate, faculty member, researcher, project or research assistant working on behalf of a faculty member or researcher.

1.1 **Determine nature (or importance) of project client is engaged in.**
The type of project is critical in deciding whether a search process will be undertaken. Is the client engaged in a project for coursework or for research? It should be noted that depending on the mandate of the library, coursework activities may rank as high in servicing the client as research activities would ordinarily. In addition, the project, even if it is for coursework, may be as important in the mind of the client as a project intended for research or publication. This is particularly the case if the client has invested a great deal of time in researching the problem. Thus, included in the definition of *importance* are the psychological aspects of the client.

1.2 **Determine time constraints.**
It is important to know under what time constraints the client operates. Is the project of limited duration, or are there plans to extend the project beyond the immediate or near future? Does the client have the time or inclination to undertake a search for information which may or may not lead to locating (and possibly acquiring) data? If the client is constrained by coursework or publication or funding agency deadlines, the client may not be able to perform a search and may prefer to redefine the research design. A student preparing a course project may more easily be able to redefine his research design, and indeed the faculty adviser may recommend that this be done. But, a researcher who has already dedicated time to preparing a research design may not want to redefine the research and therefore will be prepared to spend time locating data. It is important at this stage that the client be made aware of the often time-consuming aspect of searching for the existence of a data file which meets the design specifications.

1.3 **Determine the status of client.**
While it is difficult to establish a hard and fast rule that the staff will or will not allocate (beyond the normal amount of time) its resources for perform-

ing searches through its reference collection and outside the library, the status of the client is important in the decision-making process. As has been noted in Step 1.1, the mandate of the library will determine the level of service provided by the staff.

2.0 **Beginning the search.**
The client described in detail the type of data needed when the preliminary search through the library collection was performed (e.g., sampling frame, units of analysis, geographic location of sample, time points, variables). With this information, the library staff now proceeds to examine the reference collection. The search for data involves the library staff and the client, for it is often the case that the client brings to bear his previous experience in the field and knowledge of its literature.

2.1 **Perform library in-house search.**

2.1.1 **Examine library reference collection.**
The library staff begins by eliminating all those materials unrelated to the client's needs. If at any point in the examination a data file is located, the staff proceeds to Step 2.2.1.

2.1.1.1 **Examine codebooks and study descriptions organized by archival collection.**
Part of the reference collection at DPLS are codebooks and study descriptions which were obtained when previous requests for information were made by the library's clientele. (A codebook is the most basic piece of information that describes a data file.)

2.1.1.2 **Examine catalogs of holdings of archives and special inventories.**
The library staff examines the catalogs of holdings of various archives and special subject inventories produced by archives and other agencies. (DPLS attempts to collect the catalogs (or listings) of holdings for every known archive and special inventories which describe data concerning a single subject or topic.)

2.1.1.3 **Examine newsletters and journals.**
The library staff then turns to those newsletters and journals which communicate certain types of information.

2.1.1.4 **Examine data base directories and inventories (marginally related to client's data needs).**
If a search through information sources specifically related to the client's needs does not reveal a description of relevant data, the search is widened to include an examination of data base directories and inventories which, although generally providing very limited types of information about data collections, may provide some information on data collecting or generating agencies. Examples include Vivian Sessions's *Directory of Data Bases in the Social and Behavioral Sciences* (Science Associates/International, Inc.) and Paul Wasserman's *List 1971* [to present] (Science Associates/

International, Inc.), although DPLS's experience has indicated that these directories have never been useful for the types of searches that the staff performs on a regular basis.

2.1.1.5 **Examine miscellaneous collections of articles, monographs and other publications, and publishers' announcements.**
These materials may not be cataloged, but are shelved near the librarian's desk to be referred to infrequently. This information may be a footnote or reference in a book or article or a newspaper clipping in which a data collection was cited. In addition, the information might be notes which the data librarian made when previous searches were performed or information that was obtained during *ad hoc* conversations or attendance at lectures and conferences on data collection, quantitative research, etc.

2.2 **Determine success of search.**

2.2.1 **Complete successful search.**
If the search is successful, the library staff determines data availability (Step 3.1).

2.2.2 **Complete unsuccessful search.**
An examination of the reference collection indicates that there are no data which potentially meet the client's needs. The library staff contacts the client and requests that the client perform an additional search.

2.3 **Request client search of the literature.**
The client search is of a different nature: Because he has knowledge of different sources, such as specialized journals in his discipline and faculty members performing work in his specialty, he performs a search of relevant materials to which he has ready access outside the data library.

2.4 **Determine success of search.**

2.4.1 **Complete client successful search.**
If the client's search is successful and a citation to data is found, both client and staff proceed to Step 3.1.

2.4.2 **Complete client unsuccessful search.**
If the client has been unsuccessful in locating a data source, client and staff proceed to Step 4.1.

3.0 **Determining availability.**
Although data files may be "advertised," often there is no description of their accessibility. The library staff and client must adopt an appropriate strategy for determining whether the data are available. Will it be more appropriate for the library or for the client to transmit the information that the data are needed?

3.1 **Determine appropriateness of letter* from data library or client.**
The library staff and client determine whether it would be more appropriate for the library or for the client to write a letter to that individual, archive or agency where the data appear to reside. It is often the case, should data not reside in an archive or agency's collection, that the client, rather than the library, should correspond with the generator of the data file. This is because many generators of data are leery of placing their data in the public domain, but may agree to make their data available to an individual on an unpublicized and restricted basis (although most data libraries do provide for restricted access to data in their collection).

3.1.1 **Write letter for client.**

3.1.1.1 **Write to an archive, previous correspondence.**
For the archive with which the staff has previously corresponded there is no need to describe DPLS and its policies. The staff briefly describes who needs the data and for what purpose and asks if the data are available. Included in the letter is a request for a complete study description and a codebook, for an estimate of cost for reproducing the file, for a description of the computer specifications if the data can be provided in the needed format, and for an estimate of time required to process the DPLS request.

3.1.1.2 **Write to an individual or archive or agency, no previous correspondence.**
The library staff briefly describes the organization, its services, the data requested, and the honoring of restrictions placed upon the data. Also included in the letter is all the information contained in Step 3.1.1.1.

3.1.2 **Suggest client write letter.**
The client describes how he located the data and what his research needs are. Included in the letter may be a statement that a social science data library (DPLS) exists to facilitate data acquisition and to act as a conduit for obtaining the data file.

3.2 **Receive response.**
Because a file exists does not mean it is available. As has been noted earlier, generators of a data file are generally reluctant to place their data in the public domain.

3.2.1 **Receive positive response.**
If the data file is available and the information which the library has included in its preliminary letter has been answered satisfactorily, the library must decide whether to acquire the data. (Proceed to Step 5.0.) Data acquisition (factors influencing the decision to acquire data) is not a subject for more extensive discussion in this article, but is described in the flowchart.

*The emphasis here is on the letter medium of communication because there are many pieces of information which must be obtained. Telephoned requests are notorious for the inadequate transmission of information. Obviously, with those archives with which one deals on a regular basis, the library staff can telephone for a minimum amount of information and follow-up with an extended request letter.

3.2.2 **Receive negative response.**
If the data file is not available, the client must redefine his research project and return to the beginning of the process.

4.0 **Locating data sources.**
Success in locating resources outside the data library depends in large part on the extent and quality of the informal communications which the librarian has developed while interacting with the library's clientele and data library or archive professionals. Familiarity with archive personnel and specialists on the campus will increase the likelihood of a successful response to the librarian's query. Archives which are deluged with their own work assign a very low priority to responding to queries for information made by individuals outside their organization. Archive personnel are more likely to respond to requests from outside their organization if they know the individual requesting information. Knowledge of the specialists on the campus means that the time spent searching for information may be shortened because these specialists are aware of work through their own communications network in their area of expertise and should be able to recommend other specialists in the field.

4.1 **Perform library activities to locate potential data sources.**

4.1.1 **Write letters or telephone to archives.**
Because catalogs of holdings are dated by the time they are published, data files which are received after publication date remain unknown and inaccessible by the usual means. It is worthwhile writing or telephoning the archive to ask whether data needed by the client have been received by them.

4.1.2 **Write letters to individuals located in archives.**
Individuals in data libraries and archives often become specialists in a particular area of data collection. The data librarian writes to them, describing the client's needs and asking their assistance.

4.1.3 **Make telephone calls to specialists located on campus.**
Based on the data librarian's knowledge of faculty and research specialists located on the campus, telephone calls are made and assistance in locating the data is requested.

4.1.4 **Write letters to individuals and/or agencies recommended by specialists.**
Based on recommendations made by campus specialists, the library staff writes to these individuals for assistance, citing the campus researcher's name. The form of the letter is described in Step 3.1.1.2. (A general letter

not directed to a particular individual within an agency or institute is likely to go unanswered because the receptionist who opens the mail rarely knows to whom to direct the librarian's query.)

4.2 Suggest client write letter(s).
Upon recommendation of the specialist with whom DPLS has spoken, and/or based upon the client's knowledge of his colleagues, the client writes to potential sources of information.

4.3 Receive responses.
Responses to library and client letters and library staff telephone calls will reveal the existence and location of a data file to meet client's needs.

4.3.1 Receive positive response.
It has been determined that a data file exists. The letter received should also indicate whether the data are available. If no indication of availability is made, staff and client proceed to Step 3.1.

4.3.2 Receive negative response.
After an appropriate wait for responses (sometimes requiring that the library staff and client telephone or write a "follow up" letter requesting assistance because no response has been forthcoming) or if responses are negative as to knowledge of the existence of such data, client returns to the beginning of the process to redefine the study.

5.0 Data acquisition.
As suggested in Part I, Acquisitions Policy, criteria for data acquisition must be established. The policies must be flexible enough to meet a client's needs and to build an integrated collection of utility to the library's clientele. As noted earlier, at the beginning of the pre-acquisition process the client was informed that although a data file might be available, the library could not assure its client that acquisition would necessarily occur and that a decision to acquire a data file would be based on the quality of the data and other factors. (Data acquisition steps are included, but not detailed, in the discussion in order to complete the description of the decisions made during the process.)

5.1 Determine whether to acquire a data file.

5.1.1 Acquire data file for client (library purchase).

5.1.2 Decide not to acquire data file for client.

5.2 Suggest client acquisition.

5.2.1 Acquire data file (client purchase).

5.2.2 Decide not to acquire data file.

6.0 Data file acquisition.

Stalking the Wild Data Set: The Acquisition of Machine-Readable Social Science Data at Home and Abroad

David Nasatir

It was hot. So was I, and, apparently, so was the data set I was about to acquire. Since the scholar who had conducted the study had donated his data to Berkeley's International Data Library and Reference Service, acquisition of the IBM cards should have been straightforward. Unfortunately, the research organization storing the cards viewed them as a salable commodity—even though there was only one potential "buyer" in the world, and I was bound never to pay more than the costs of duplication. By general agreement, "ownership" resided with the scholar. But possession, as they say, "counts for a lot." So I requested the taxi to wait.

The heat provided a compelling reason for the siesta after lunch, and I counted on the absence of the higher level officials with whom I had spent the morning arguing. The driver was convinced of my insanity when I emerged with the first case of cards (about 20 kilos worth), and shook his head in disbelief as I returned with the second and third loads. The coat, tie, heavy loads and sweat stains were incongruous, to say the least.

Off we sped to the IBM office where, by prior arrangement, a staff member was standing ready to copy the cards to tape. The driver's disbelief was audible when, shortly after I had started carting the cards into the office, I began to cart them back out again and, finally, ordered that we return to our first destination. The caper was completed before the higher-ups returned, and I suppose they are still wondering how that particular data set eventually appeared in various catalogs of studies around the world. But acquisition isn't always so easy.

Acquisition for any special library is a difficult and challenging task, but difficulties may be compounded when the materials consist of machine-readable data, and the challenge is often increased when the data refer to sensitive topics. The sensitivity of the topic may change, of course, and materials that were initially beyond hope of acquisition may become available. The reverse is also true, and a great deal of effort can be devoted to identifying materials worthy of acquisition only to have international relations become strained while en route to acquire them.