Nonconventional Data Sources and Reference Tools for Social Science and Humanities

This essay attempts to provide an introduction and guide to the emerging literature on nonconventional information systems and centers which serve social science and humanities in the United States. Coverage includes the changing bibliographic and data needs; the centers that are offering nonconventional, library-type services; the services offered; the coordinating organizations; computer-compiled reference tools; and other sources of information for librarians assisting scholars who wish to employ new approaches to research in the social sciences and humanities. Included is a list of publications and articles which in combination may serve as a directory.

Innovations in scholarship in the social sciences and humanities are reflected in published studies which are being added to library collections across the country. The new approaches to research are made possible by the application of computer technology to the analysis of quantitative data, to information processing, and to the study of literary style. Thomas J. Condon, an executive associate of the American Council of Learned Societies, expressed the possibilities of some of the nonconventional techniques when he said:

The computer holds out . . . the promise of providing the scholar in the humanistic and social science disciplines with the kind of information, the quality of information, and the access to information that he scarcely ever permitted himself to imagine possible. From the level of activity that could well develop if the scholar and the machine can work in tandem may come a closer approximation of the ideal of a unified culture than we ever managed before.

Effective reference librarians and research assistants have long had at their disposal directories listing sources of specialized information. Referral of the user to a special collection held in another library or to other sources of needed information has shortened the search for many a scholar. In order to fulfill their usual role as guides in the search for available information it is essential.

1 Examples of recently published books which have resulted from new trends in scholarship are the following:

Mrs. Gulick resides in Potsdam, New York.

Nonconventional Data Sources and Reference Tools / 225

that today’s social science and humanities librarians have knowledge of the nonconventional as well as the conventional sources.

Nearly all librarians are now using reference tools that are compiled with the assistance of computers. For more than twenty years science librarians have been aware of the possibility of literature searches done at special centers which employ extensive subject indexing of their collections. The methods of data and bibliographic control which were first developed to serve science have only recently been employed to serve social science and humanities. Financial support (foundation, business, and government) for their continued development is increasing.3

This article, with the references which follow, will serve as an introduction and guide to the emerging literature on nonconventional information systems and centers which serve social science and humanities in the United States. Librarians and subject specialists exploring these possibilities for the first time will ask such questions as:

What research is being done toward meeting the bibliographic and data needs of the social sciences and the humanities?

Where are the centers that are now offering nonconventional, library-type services?

What are the services offered?

What organizations are concerned with coordinating the services of the developing network of centers?

What are the advantages of the new bibliographic tools which have been made possible by the use of nonconventional techniques?

DEFINITIONS

The terms “system” and “center” require definition. The distinction can be made clear by first defining information retrieval as the total process of gathering, analyzing, indexing, filing, and making available items of recorded information. An information retrieval system is an organized combination of the elements (human and/or hardware) especially designed to carry out the information retrieval process.4 The center is the place where the process is carried out.5 A “nonconventional” system is one which uses such searching devices as Uniterm cards; interior or edge-notched punched cards; standard punched cards; punched paper tape; magnetic cards, tape, discs, or drums; or unit or serial microform for the storage of document indexes. The system is also “nonconventional” if unique classification, indexing, or coding techniques not commonly associated with usual library practices are employed.6 The purpose of a center using a nonconventional system may be for (1) the storage of references; (2) the storage of data or data compilations; (3) the production of bibliographies and indexes; or (4) any combination of these three.

Information which is the raw material of the social scientist is of two types: source data (statistical reports, survey results, voting records, etc.) and the literature or the texts that have resulted from the analysis of the source data. The

3 For consideration of the bibliographic and data needs of the social sciences and humanities, see:

Eric H. Boehm, Blueprint for Bibliography: A System for the Social Sciences and Humanities (Santa Barbara, Cal.: Clio Press, 1965).

Eric H. Boehm, “Dissemination of Knowledge in the Humanities and Social Sciences,” ACLS Newsletter, XIV (May 1963), 3-12.


storage of the source data through the use of nonconventional techniques creates what are now termed "data banks" or "data archives." Storage may require reduction of a natural language text to a standard card size or film, or it may involve coding for storage on punched cards, magnetic tapes, discs, or drums. Bibliographic control of the literature of the social sciences may be accomplished through the use of techniques which permit rapid compilation. The process is one of storing an extensive index which can be searched by subject, author, title, geographic unit, or any other aspect deemed relevant and entered on the record.

The raw material of the literary scholar and critic is the literary text. Computer-compiled analytical indexes and concordances have become familiar tools. The first step in using data processing techniques for text analysis is the entering of the text on magnetic tape or disc via punched cards. Once the text is so recorded, it can be copied onto other cards, discs, or tapes. Computer programs for various kinds of analysis as well as the recorded texts can be useful to scholars other than the ones for whom they were first developed—if the accessibility of the coded texts and programs is known. As in the case of the social sciences, electronic information storage and retrieval systems can be used for bibliographic control of the literature of the humanities.

EXISTING CENTERS AND COORDINATING ORGANIZATIONS

With the support of the National Science Foundation and the collaboration of existing data archives, the Council of Social Science Data Archives was established in 1965. Ralph L. Bisco, technical coordinator for the organization, describes it as a planning, policy making, and information-disseminating group for coordinating and publicizing the activities of a confederation of social science data archives, most of which are in the United States. Its basic purpose is to make machine-readable data and supporting documentation readily accessible, at minimum cost, to scholars. The concern of the Council of Social Science Data Archives extends only to those archives that routinely provide machine-readable data to users outside the original research group or data collection agency. One of the first projects of the Council has been the compilation of a directory which describes the holdings of twenty-five such archives, Social Science Data Archives in the United States—1967. Further information about the Council and its member archives is available upon request from the Executive Director, William A. Glaser, 605 West 115 Street, New York, New York.

Archives which provide machine-readable data to users outside the collection agency are of two types. One is the general purpose service archive which makes its holdings available to the entire community of social scientists; the other is the local service archive which provides data for a more limited group of users.

One of the most influential general purpose service archives in the United States is the Inter-University Consortium for Political Research. It is a nonprofit partnership between the Survey Research Center of the University of Michigan and the National Science Foundation. With the support of the National Science Foundation and the collaboration of existing data archives, the Council of Social Science Data Archives was established in 1965. Ralph L. Bisco, technical coordinator for the organization, describes it as a planning, policy making, and information-disseminating group for coordinating and publicizing the activities of a confederation of social science data archives, most of which are in the United States. Its basic purpose is to make machine-readable data and supporting documentation readily accessible, at minimum cost, to scholars. The concern of the Council of Social Science Data Archives extends only to those archives that routinely provide machine-readable data to users outside the original research group or data collection agency. One of the first projects of the Council has been the compilation of a directory which describes the holdings of twenty-five such archives, Social Science Data Archives in the United States—1967. Further information about the Council and its member archives is available upon request from the Executive Director, William A. Glaser, 605 West 115 Street, New York, New York.

Archives which provide machine-readable data to users outside the collection agency are of two types. One is the general purpose service archive which makes its holdings available to the entire community of social scientists; the other is the local service archive which provides data for a more limited group of users.

One of the most influential general purpose service archives in the United States is the Inter-University Consortium for Political Research. It is a nonprofit partnership between the Survey Research Center of the University of Michi-
gan and over ninety universities and nonprofit research organizations in Canada, Europe, Great Britain, and the United States. The collection includes most of the major political survey research of the past thirty years, the 1950 and 1960 censuses, and major election statistics and population characteristics, by county, from 1820 to the present. In 1964 the collection was described as including ten thousand interview forms of unevaluated data, eighty reels of tapes, twenty linear feet of project cards (collected at the rate of fifteen thousand cards a year), and four linear feet of technical reports. The basic storage medium has been magnetic tape. Data services offered by the Inter-University Consortium range from a simple reproduction of card and tape files of complete studies, through the preparation of special analysis or work decks or tapes tailored to the analysis needs of particular researchers, to the provision of percentaged tables and multi-variate analyses of data.

In addition to the Inter-University Consortium for Political Research, examples of the general purpose service archives are the Roper Public Opinion Research Center at Williams College, the International Data Library and Reference Service at the Survey Research Center of the University of California at Berkeley, the Bureau of Labor Statistics, the Bureau of the Census, and the Louis Harris Political Data Center at the University of North Carolina. Examples of the local service archives are the Center for International Studies at Massachusetts Institute of Technology and the Political Science Research Library at Yale University.

Other collections of social science data have been organized in nonconventional form although their files are not machine-readable. Such files are usually maintained on cards or film and use unique classification systems which are recorded on cards that can be sorted manually or mechanically. The most venerable examples are the Human Relations Area Files and the Classified Abstract Archive of the Alcohol Literature. A careful perusal of the descriptions of the special libraries listed in a 1965 publication of the National Referral Center for Science and Technology, A Directory of Information Resources in the United States, Social Sciences, will reveal descriptions of fourteen centers using nonconventional systems of data storage. Half of the nonconventional systems listed do not maintain machine-readable files. Some offer to perform searches upon request; other supply duplicate sets of punched cards. Nearly all permit academic researchers to apply for access to the collections.


The Bureau of the Census Catalog, issued quarterly and cumulated annually, lists in addition to the publications issued by the Bureau unpublished material including data files, selected special tabulations, and nonstatistical materials such as maps and computer programs. In addition the Bureau of the Census in 1967 established the Data Access and Use Laboratory which edits Small-Area Data Activities, a newsletter issued occasionally for the purpose of informing readers of developments in the field of small-area data, and Data Access Descriptions, a publication which covers specialized aspects of access to unpublished census data. The first two issues of the latter were published in March and August of 1967. Additional information may be requested from John C. Beresford, Office of the Director, or from Mrs. Constance Citro, Data Access and Use Laboratory, Bureau of the Census.


Mark Keller, "Documentation of the Alcohol Literature: A Scheme for an Interdisciplinary Field of Study," Quarterly Journal of Studies on Alcohol, XXV (December 1964), 725-41.
Since 1964 the American Council of Learned Societies has offered grants-in-aid to scholars in humanities and social sciences in order to encourage computer applications in their research. In June of 1966 a special supplement of the ACLS Newsletter listed 263 social science and humanities projects involving the use of electronic computers in progress in the United States.

The fields of archeology, anthropology, history, language, literature, and music were most represented, although a few projects were underway in such diverse fields as religion, psychology, and architecture. Publication of the list provided a means of communication among scholars of similar interest. The American Council of Learned Societies has more recently undertaken the Center’s Journal Editors project. This is a pilot project to test the feasibility of interdisciplinary bibliographic control through computerization. Thirty journals published by the constituent societies of the ACLS are involved. A merged quarterly index which is available simultaneously with the publication of individual journals is anticipated. The Center, while accumulating a data bank, will develop techniques of selective dissemination of information for interdisciplinary work in the humanities and social sciences, produce specialized indexes reflecting interdisciplinary concerns, and prepare an annual merged index of constituent society journals.

The United States Office of Education created the Educational Resources Information Center (ERIC) program in 1964 as a means of increasing the circulation and use of research reports and research-related materials covering education at all levels. The present eighteen ERIC centers for acquisition have been located where outstanding collections on the various fields of education already existed. At present, report coverage is available on the following subjects: counseling and guidance, the disadvantaged, educational administration, exceptional children, foreign languages, junior colleges, linguistics, reading, school personnel, science education, small schools, vocational education, adult and continuing education, teaching of English, library and information sciences, educational media and technology, educational facilities, and early childhood education.

At the acquisition centers the abstract of each document and the index terms which reveal its total content are combined to form the résumé. The original report is then copied on microfiche for deposit in a centralized reproducing and distributing center in Cleveland. Announcement bulletins entitled Research in Education are published monthly. The bulletins include the abstracts and list author, location of study, and length. Requests are made by the users to the distribution center which will supply either microfiche or hard copy text of the complete report.

Under the present arrangement, each ERIC clearinghouse performs two services in addition to selection, abstracting, and indexing. Each center responds to inquiries from the educational community which cannot be handled with ERIC’s routine services and prepares special reports such as critical reviews, selected bibliographies, and state-of-the-art papers as they are needed.

The Urban Renewal Administration of the United States Department of Housing and Urban Development in cooperation with the New York City Planning Commission and the City Universi-
The Interuniversity Communications Council (EDUCOM), an association of over sixty institutions of higher education, was created in 1965 chiefly for the purpose of informing staff members of the member institutions about new technologies which can be used in education. Nonconventional techniques for storage and retrieval of information and computational and linguistics uses of computers will continue to receive EDUCOM encouragement.18

Fifteen organizations and agencies dealing with information handling in the social sciences were represented at a conference sponsored by The Galton Institute in October of 1966. The meeting has been described as the first step toward cooperative resolution of difficulties encountered in the processing, storing, disseminating, and retrieving of documents and data in the fields of mutual interest.19 As an outgrowth of this meeting a new division of the American Society for Information Science (formerly the American Documentation Institute) was created. It is called the "Special Interest Group in the Behavioral


Sawin and Charles Nilon in 1958. A pilot study for testing the feasibility of the proposal has now been completed at the University of Colorado under a grant from the U.S. Department of Health, Education, and Welfare. The journal coverage of the proposal, the arguments presented against its practicality, and now the report of the pilot study provide a good introduction for the librarian or library school student who is considering the new techniques for the first time.

Some early examples of KWIC indexes in the social sciences and humanities were Kenneth Janda's *Cumulative Index to the American Political Science Review*, Vols. 1-57, 1906-1963 (Evanston, Illinois: Northwestern University Press, 1964) and the *Kansas Slavic Index: Current Titles: Social Sciences, Humanities* (Lawrence, Kansas: University of Kansas Libraries, 1963). Comparative Education Index and Bibliography was first produced at the University of Michigan in 1964. As early as 1964 and 1965 indexes to *PMLA* were done by electronic data processing; the Modern Language Association expects to have a completely automated bibliographic system in operation after 1968.

The first issue of *RILM Abstracts of Music Literature* was published in August 1967. It is an abstracted, computer-indexed bibliography of current scholarly literature on music. Eventually RILM (Répertoire International de la Littérature Musicale) will undertake publication of a series for retrospective material. The project is jointly sponsored by the International Musiological Society and the International Association of Music Libraries. It has received support from the American Council of Learned Societies and is serving as a pilot project for that organization’s proposed interdisciplinary National Bibliographic Center. The project is directed by Professor Barry S. Brook, Queens College, The City University of New York.

The first of a proposed ten-volume series of bibliographies and indexes on political science, government, and public policy was issued by the Universal Reference System in 1965 (*International Affairs. Universal Reference System, Political Science, Government and Public Policy Series*. Vol. I.). Nine additional volumes and an updated version of the original volume should be available in 1968. The method of production employs a unique classification system and intensive indexing as well as computerized handling of the data involved. The published volume is called a “codex” instead of an index because most of the descriptors have been truncated in code words and the descriptors have the inherent ability to be crossed with any set of descriptors. The codex has two main parts, a catalog of annotated titles and a much larger index-in-depth. Each document in the bibliography may have from ten to twenty entries in the index. Coverage includes books, magazines, and other forms of documents ordinarily listed separately. Another departure

---

from tradition is that the descriptors used for indexing not only cover the subjects usually found in library subject cataloging but are also used to reveal previously unindexed aspects of documents such as the ethical base of the authors and the methodological techniques used in the research. When the complete system is available, each of the ten fields of political science which will be covered by a codex will also be covered by a quarterly gazette which can be used to keep the scholar abreast of current published material. Each quarterly gazette will cumulate the previous issue up through the first year of the cycle. An updated codex in each of the ten fields is proposed for publication every two years. Gazettes are currently available in seven of the ten fields. The service to be offered, a computerized documentation and information retrieval system, will be year-round on a two-year cycle. Further information is available from Princeton Research Publishing Company, 32 Nassau Street, Princeton, New Jersey.

The interdisciplinary approach to social science research has resulted in a need on the part of the scholar for access to literature outside his field of specialization. One of the more recent index journals which attempts to meet this need for those working in the behavioral-biological-social science field is *Perceptual-Cognitive Development*. Each issue includes a computer-compiled bibliography, a keyword-in-context index, and an author list. Its broad coverage and the possibility of listing an article within two to five months after publication are given as its chief reasons for existence.

The citation index is another innovation resulting from the use of computers. Because compilation of a citation index requires a minimum of human intellectual effort, it can be available to users sooner after publication of the books and journals covered than can conventional indexes or even computer-compiled indexes which are based upon abstracting. With the present extended coverage of *Science Citation Index*, Eugene Garfield, its creator, is proposing its use by social scientists who want to "keep up" with related scientific fields. Within the next few years the Institute for Scientific Information expects to extend its coverage to more than three hundred key journals in such fields as sociology, political science, anthropology, history, and other areas of the behavioral sciences. A personalized current awareness service of the type now offered for scientists can then be available for social scientists.

**DIRECTORIES**

No single directory exists which can be used as a guide to the sources of social science and humanities information held in centers which employ non-conventional storage systems. Likewise, no single directory of computer programs for data processing exists. The following publications and articles indicate developments to date and for the present they may serve the function of directories.


*Published since 1965 by the Galton Institute, P.O. Box 35336, Preuss Station, Los Angeles, Cal. 90035.*

---


29 Published since 1965 by the Galton Institute, P.O. Box 35336, Preuss Station, Los Angeles, Cal. 90035.

Carlson, Gary (comp.). "Literary Works in Machine-Readable Form," Computers and the Humanities, I (January 1967), 75-102. (A list of 224 literary works in machine-readable form and the names and address of 66 researchers with whom interested persons might correspond.)

"Computer Programs Designed to Solve Humanistic Problems," Computers and the Humanities, I (November 1966), 39-55. (Lists 36 programs for text manipulation, 8 for music, 2 for graphic display, 4 for social science—all with descriptions of purpose and name and address of developers.)

"Additional Programs to Solve Humanistic Problems," Computers and the Humanities, I (January 1967), 108. (A list of eight programs not included in the previous issue.)


Basic Books

Librarians and subject specialists who for the first time are investigating the techniques and potentials of nonconventional information storage and retrieval as well as computer applications in the social sciences and humanities may find the following volumes of interest.


²⁹ Of the 175 nonconventional information systems serving science and technology which are described in this volume, at least four maintain collections which are of interest to social scientists. They are the International Data Library and Reference Service at the University of California at Berkeley, the U.S. Atomic Energy Commission's Serial Record System, the National Institute of Mental Health's National Clearinghouse for Mental Health Information, and the National Bureau of Standards' Clearinghouse for Federal Scientific and Technical Information.
Nonconventional Data Sources and Reference Tools / 233


SELECTED PERIODICALS

The total number of journals which occasionally carry articles of interest to persons wishing to be continually informed on this interdisciplinary topic is beyond the scanning capacity of the average person. It is possible, however, to watch the regular publications of a few journals which give better coverage to the topic than do others. The following nine are suggested:

Computing Reviews. New York: Association for Computing Machinery, 1960-


United States Government Research & Development Reports. Washington: United States Department of Commerce, 1938-


EDUCOM: Bulletin of the Interuniversity Communications Council. Pittsburgh: Interuniversity Communications Council, 1966-

International Social Science Journal. New York: UNESCO Publications Center, 1949-

AGLS Newsletter. New York: American Council of Learned Societies, 1949-

Social Science Information. Paris: International Social Science Council, 1962-

OTHER SOURCES

A good source for current information from the moving edge of any field that enjoys governmental support of its research is the government report. Information contained in journals today may have been available as government reports a year ago. On the other hand, some valuable information may never become available in any form except that of the government report. Hence, the anonymity of the unsuccessful research project. The Government-wide Index to Federal Research and Development Reports (Washington: U.S. Department of

Volume X, Nos. 5 and 6 (January and February, 1967) were devoted to “Advances in Information Retrieval in the Social Sciences.”

Published with financial assistance from the IBM Corporation and the U.S. Steel Foundation. Address inquiries to Prof. Joseph Raben, Department of English, Queens College, Flushing, New York 11367.

This publication is being distributed without charge, experimentally, among the faculty at member institutions. Membership is open to all accredited colleges and universities in the United States, Canada, and Mexico. Address inquiries to Carle Hodge, Editor, 4200 Fifth Avenue, Pittsburgh, Pennsylvania 15213.

Much of Volume XVI, No. 1 (1964) was devoted to “Data in Comparative Research.”

Published monthly (October-May); available without charge. Address inquiries to the American Council of Learned Societies, 345 E. 48th Street, New York, New York.

In the future a special section concerning data archives will be regularly included in this publication. Volume IV, No. 3 (September 1965) was devoted to data production, storage, retrieval, and processing.
Commerce, 1965- ) serves as an index to most unclassified reports on federally financed research. Through the Fast Announcement service of the Clearinghouse for Federal Scientific and Technical Information persons interested in specific fields can be assured of notification as the reports become available.

Not to be overlooked are the sources from which information can be obtained upon direct request. Unpublished papers which were presented at meetings and periodic reports of organizations involved in research and development are important sources of information relevant to nonconventional, library-type services for social science and humanities. 29

29 Examples of relevant unpublished papers are: