

## ONLINE SEARCHES OF SOCIAL SCIENCE DATA SETS:

### THE RIQS SYSTEM AND ICPSR DATA

Ann Janda and Kenneth Janda  
Northwestern University

Every solution seems to generate a new problem. The problem of accurately assessing public opinion led to the invention of the sample survey. The subsequent problem of analyzing survey responses brought widespread use of machine-readable data. The problem of preserving machine-readable data for secondary analysis stimulated the creation of data depositories or "archives." Growth over time in the holdings of these social science data archives, however, has aroused needs for improved retrieval of data. This paper explains one method of dealing with such needs. It involves an interactive search of the holdings of the most diversified social science data archive, the Inter-University Consortium for Political and Social Research, using a general-purpose information retrieval system, RIQS, written for CDC computers.

#### Nature of the Problem

The growth of social science data archives has been chronicled by Sidel (1980). The "golden decade" of data archives was the 1960s, when social science data depositories sprang up at universities across the United States and Europe. As early as 1967, efforts to coordinate archiving activities in the U.S. led to formation of the Council of Social Science Data Archives, a cooperative federation of 25 member organizations (Bisco, 1967). Among the Council's main objectives were efforts to prepare a directory of data files archived by member institutions, to standardize study documentation, and to develop a retrieval system to access the holdings. With the loss of federal funding in 1970, the CSSDA ended, its objectives largely unfulfilled (Sidel, 1980: 54).

The rapid growth of data archives in the 1960s would have suggested their proliferation in the 1970s; but that did not occur. Hofferbert and Clubb found that "contrary to expectation, growth has been

vertically, in terms of size of holdings and level of services, rather than horizontally, in terms of number of archives" (1976: 382). The archive that grew the most "vertically" since its founding in 1962 was the Inter-University Consortium for Political Research, renamed the Consortium for Political and Social Research in 1975. With its membership recently reported to be 237 institutions and national affiliates (Annual Report, 1979-1980: 125), the ICPSR has emerged as the dominant data archive in the social sciences. In this capacity, however, the Consortium has done more than act as a data archive. It has undertaken major data collections on its own, sponsored a highly successful continuing summer program offering courses on methodological and analytical topics, and pioneered in development of computer applications to social research. Its greatest impact on students and scholars across the country, however, has come through its role in archiving and disseminating social data for secondary analysis.

The value of secondary data for scholarly research was noted early by Rokkan (1964), and its value for teaching has been emphasized recently by Sobal (1980). Unfortunately, carefully preserving data in an archive does not itself insure that the data will be used effectively in teaching or research. Dodd notes that data in archives "are not fully utilized because potential users are unaware of the existence and accessibility of social science data studies" (1977: 48). As the number of studies in an archive increases, the very richness of its holdings complicates the problem of retrieving the data. One must wade through numerous study descriptions or abstracts, which in themselves may not adequately describe the contents of the study. In principle, this problem might be resolved with the help of knowledgeable data librarians. Conger suggests that "the main satisfaction of reference" work for librarians working with machine-readable data files in the social sciences stems from gaining familiarity with the variables contained in the studies. Armed with such expertise, the librarian might direct a student interested in women's liberation to a seemingly unrelated dataset, like the National Longitudinal Survey of Labor Market Experience (1976: 1964). Unfortunately, there are not enough skilled reference librarians with such knowledge to go around. Instead, a powerful, automated means to retrieve data from a large data archive could serve as well. It would allow the

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

user to peruse the holdings according to his research interests independent of a reference person, a factor which Tanenbaum cites as an important feature of an archive management system (1979: 429). Indeed, Peters contends that the "top priority" be given to the development of an information retrieval system for social science data rather than to a dissemination system (1974-1975: 8).

#### Features of a Retrieval System

The "complete" data archive and "ultimate" retrieval system are more easily imagined than realized. Imaginations soared during the early part of the golden 1960s, when participants in national conferences on data acquisition, storage, and retrieval shared their visions of the future. One speaker described data banks so rich in information that we could compute even "the half-life of a prejudice in food" and the "military participation ratio in the crusades" (Deutsch, 1970: 35-36). We learned soon enough, however, that it was easier to visualize than to create the complete data archive. We found it difficult enough to collect complete returns for contemporary elections without archiving data on the Middle Ages. More recently, in a similar vein, Dodd constructed a scenario of the "ultimate" retrieval system that enabled one "Professor Orrie Wells" to do several months' research (by today's standards) in an hour or so at tomorrow's library (1978). While it is stimulating to imagine the ideal retrieval system, we find our energies consumed in developing modest systems to achieve limited objectives.

In truth, there have been few major advances in data retrieval since the mid-1960s. White's recent Reader in Machine-Readable Social Science Data (1977) relies on articles published in 1967 and 1968 in the section on "Indexing and Cataloging Social Science Data." A systematic search of entries under "Behavioral and Social Sciences" and "Information Retrieval" in Computing Reviews discloses no published reports of improved data retrieval systems in operation. But the field is not entirely without movement. Accounts have been written by Dodd, describing the system serving the social sciences through the Triangle Universities Computation Center Network Environment (1978); by Ferguson, explaining the application of the SPIRES retrieval system to data holdings at Stanford (1977); and by Tannenbaum, writing about the data management system at Essex (1979). Understandably, each of these systems is tailored to local computing environments and to local dataset documentation. Unfortunately, these local peculiarities interfere with exporting these systems to other archives.

A major stumbling-block in building a successful information retrieval system lies in preparing the input to be retrieved. Sidel writes, "The difficulties encountered in developing inventories of data sets and computer programs all stem, at least in part, from lack of standards for the form of data holdings and the documentation maintained and distributed" (1980: 55). In addition to the formatting issue, there is the mundane problem of keystroking the study and data descriptions. It would be much easier to develop an

exportable data retrieval system if we had "standardized" documentation already available in machine-readable form. In fact, we have documentation that fits this description rather nicely, although few have recognized it.

The Consortium furnishes two forms of machine-readable text suitable for use in a data management system. The ICPSR publication, Guide to Resources and Services, which contains the descriptions of all studies in the archives, is available from the Consortium on magnetic tape, as are some study codebooks, particularly those for national sample surveys. Due to the widespread membership in the Consortium and the breadth of its holdings, this machine-readable documentation provides a common and important core for the management system of any social science data archive. Because the Northwestern data management system is designed to accept and search Consortium documentation, the idea should be exportable, at least in principle, to any other institution (1) that has Consortium membership and (2) that possesses general-purpose information retrieval software.

#### The ICPSR Holdings

Although the ICPSR does not conveniently tally the studies in its holdings, a careful count of study descriptions in the ICPSR Guide to Resources and Services, 1980-1981 shows 718 studies available to Consortium members for secondary analysis. While this number seems small compared to over 15,000 polls contained in the Roper Public Opinion Research Center (White, 1977: 131), one must realize that the ICPSR count includes multiple studies and much larger files, many of which are not surveys, in a wide variety of social science fields. The distribution of ICPSR Archival Holdings by 17 categories in the Guide to Resources and Services, 1980-81 is reported in Table 1.

The ICPSR was originally formed to archive the national surveys of the University of Michigan's Survey Research Center. Its holdings, which were initially concentrated in studies of political and economic attitudes, have become diversified over time. The two categories of "Survey Studies of Mass Political Behavior and Attitudes" and "Economic Behavior and Attitudes", together now account for only about 40 percent of the studies in the ICPSR archives. The remaining bulk of the data sets are distributed over many broad categories with the greatest number in "Social Institutions and Behavior." Clearly, students and scholars in institutions that are members of the Consortium have potential access to a wide range of data.

"Potential" access is the key word. Assume that the prospective user is directed to the local Consortium representative for information about its holdings. How does this student or scholar go about searching for relevant data? In large part, this depends on the institutional commitment to Consortium membership and the staff support available to help the new user. But in all cases, the starting point is likely to be the annual ICPSR Guide to Resources and Services, especially the 17 main categories of "archival holdings" and their subdivisions. On consulting this classification,

TABLE 1: Distribution of ICPSR Archival Holdings by Categories for 718 Studies in 1980-81

	Number	Percent
I. Census Enumerations: Historical and Contemporary Population Characteristics	35	5%
II. Community and Urban Studies	38	5
III. Conflict, Aggression, Violence, Wars	41	6
IV. Economic Behavior and Attitudes	93	13
V. Education	21	3
VI. Elites and Leadership	22	3
VII. Environment and Natural Resources	1	0
VIII. Government Structures, Policies, and Capabilities	30	4
IX. Health Care and Health Facilities	6	1
X. Instructional Packages and Computer Programs	26	4
XI. International System: Linkages, Relationships, and Events	47	7
XII. Legal Systems	12	2
XIII. Legislative and Deliberative Bodies	25	3
XIV. Mass Political Behavior and Attitudes	164	23
XV. Organizational Behavior	15	2
XVI. Social Indicators	45	6
XVII. Social Institutions and Behavior	97	14
	718	101

<sup>a</sup>A simple count of studies in the Guide to Resources and Services, 1980-1981 (Inter-University Consortium for Political and Social Research),

users are directed to the pages that contain the descriptions of these studies. Some lucky ones may find their needs satisfied by browsing through the study descriptions under the relevant headings. Most users, however, find that their interests overlap one or more the headings and become overwhelmed by the number of descriptions to be read. Additional help comes from three indexes to the studies in the back of the Guide, where the studies are listed in alphabetical order (a) by title, (b) by principal investigator, and (c) by subject.

Most new users will find the subject index the most helpful. The index in the 1980-81 Guide runs 29 pages, using terms ranging from "abortion" to "Zambia." Despite its length, this index, which is not computer-generated, is incomplete. For example, a careful count of the studies listed under "cross-national" in the 1979-1980 Guide suggests that the entry misses about half the cross-national studies (Janda, 1980). Even with a perfect index to the study descriptions, the Guide falls short as a retrieval device because it is limited to describing the studies and not the variables in the studies.

To be effective, a data retrieval system should operate at two levels: the data file and the data item (Nielsen, 1974-1975, and Mochmann, 1974-1975). Nielsen defines the data file level to include standard catalog entries for a study, study abstracts or descriptions, and physical characteris-

tics of the file (p.40). Mochmann includes in the data item level all information pertaining to the specific variables in the study: the questions producing the response in the case of survey data, the response categories, variable names, and other codebook information (p.48). A search of information on data files should identify studies that bear more or less directly on the user's interests. A search on data items should locate variables that pertain more specifically to the user's needs even though they may be tucked away in ostensibly unrelated studies. The ultimate retrieval system would contain information at both levels and matching search capabilities, but this is more easily conceived than realized. We have attempted to move in the direction of a two-level search of the ICPSR holdings in our current online data management system at Northwestern, using only the machine-readable information distributed by the Consortium and a general-purpose information retrieval system.

#### The RIQS System

RIQS stands for Remote Information Query System, a general-purpose computer program for information storage and retrieval that was developed at Northwestern University in 1970 (Mittman and Borman, 1975: 34-36). The program was written for the Control Data Corporation 6000 computer series. Besides Northwestern University, it has been

implemented at 21 CDC installations--12 of these outside the United States. Although about two-thirds of the RIQS program is in FORTRAN, the large word size of CDC computers severely limits RIQS conversion to other computers. Moreover, all the input-output routines are in CDC COMPASS, even further restricting the exportability of RIQS to non-CDC installations. Fortunately, the RIQS program is not essential to the data management idea to be discussed; its limited exportability should not prevent implementing a similar data retrieval system elsewhere using a different information retrieval program.

For our purposes, we need outline only the essential features of the RIQS system for creating and searching files. A RIQS file is composed of a series of records that consist of user-defined items of information. In our application, we created two RIQS files: one containing every study in the Guide and a separate "experimental" file for selected studies with machine-readable codebooks. The first file, called GUIDE, was originally devised and created by Lorraine Borman, Manager of Information Services at Northwestern's Vogelback Computing Center. In GUIDE, each study is a RIQS record defined by these items of information:

1. Study Author
2. Study Title
3. ICPSR Study Number
4. ICPSR Archival Category and Code Number
5. Northwestern Data Library Access Information
6. Northwestern Documentation Available
7. ICPSR Study Abstract

The information stored in RIQS items 1, 2, 3, and 7 (ICPSR Study Author, Title, Number, and Abstract) is supplied by Consortium tapes containing the Guide and quarterly updates of new acquisitions. A short program was written to read the tapes and convert the information for input to RIQS, leaving only items 4, 5, and 6 (ICPSR Category/Code, and Northwestern access and documentation information) to be added separately. The current GUIDE file has 791 records (studies). This number is larger than the count of 718 studies from the 1980-1981 Guide for two reasons: the RIQS GUIDE file contains all the updates distributed through January, 1981 (which post-dates the ICPSR Guide publication), and also treats the APSA SETUPS as 21 separate studies.

The second RIQS file is called VARLIST. Each record in VARLIST has the same first seven items of information and as many additional items as necessary to handle the entries from the "variable description lists" in the Consortium's machine-readable codebooks. Although RIQS is limited to 500 items of information, each item may contain 150 subitems. Our procedure for handling the numerous variables contained in the ICPSR codebooks was to

assign approximately the first 100 variables to item 8 (treating them as 100 subitems), the next 100 variables to item 9, and so on. This procedure created a small VARLIST file when measured by the number of records (studies) but a large file when measured by the number of items and subitems (variables). VARLIST, then was used for searching at the data item level, while GUIDE was used for searching at the data file level.

RIQS can be directed to search text in individual items for the occurrence of specific words or for combinations of words according to Boolean logic. If it finds words in the particular item or items that satisfy the search command, RIQS can print the entire source record, or only specific items of the record as directed. This capability within RIQS enables the file of ICPSR studies to be searched for keywords that occur in the entire study record and to print its citation information as well as the entire abstract, if desired.

Experienced RIQS users would have no difficulty writing commands to search the GUIDE file to locate relevant studies. Our goal, however, was to offer a user-oriented system that required no knowledge of RIQS, merely an interest in Consortium holdings and familiarity with remote terminals. We decided to create a basic procedure that used only a fraction of RIQS' search capabilities and presented the user with a few simple options. This procedure, called SEEKER, operates with prompts in an interactive mode. Even neophytes find SEEKER quite easy to use by referring to a one-page handout reproduced as Figure 1. After logging in, the user simultaneously attaches the RIQS system and the GUIDE file by typing

```
%$CALL,SEEKER.
```

This command calls in RIQSONL, the online version of RIQS, with its standard prompts (not all of which are necessary to SEEKER and which await elimination). The user must then decide between searching with AND logic or OR logic by selecting the FINDAND or FINDOR procedures. The user must also decide whether to print full abstracts of any studies retrieved (YES or NO) and must specify three keywords (in place of KEY1 KEY2 KEY3) for use in the search. These are the only entries the user must make. In the following example, SEEKER will be illustrated by application to a hypothetical retrieval problem facing a potential user.

#### Retrieving Data on Abortion Using SEEKER

Suppose a student sought to analyze attitudes toward abortion using Consortium data. The 1980-1981 Guide to Resources and Services lists 21 studies in its archival holdings on "Public Opinion on Political Matters" in the United States and another 97 under "Social Institutions and Behavior." Confronted by these numbers, the student could turn to the subject index, which lists only four studies under the heading, "Abortion." Using SEEKER, the student could search the study abstracts to locate additional relevant studies. The results from the SEEKER search for ABORTION, ABORTIONS, and

Three commands are used to retrieve author, title, ICPSR No., study abstract, N.U. access and codebook information:

1. / %\$CALL, SEEKER. ----- This command attaches the database and calls in RIQSONL. Answer the RIQSONL prompts and proceed to step II.
- ? RETRIEVE { FINDAND FINDAND } ----- Select one of the two options:  
           -OR-  
           { FINDOR FINDOR }  
 A. FINDAND uses AND logic among keywords and results in a more specific search with fewer citations retrieved.  
 B. FINDOR uses OR relationships among keywords and retrieves the greater number of citations.
- ? CALL { FINDAND } ( ('YES' 'KEY1' 'KEY2' 'KEY3') END) ----- Options:  
           -OR-  
           { FINDOR } ( ('NO' ) )  
 A. The same option selected in step II must be repeated.  
 B. Within parentheses four parameters must be specified in quotes.  
 1. 'YES' will print study abstracts; 'NO' suppresses this option  
 2. Three search terms or keys must be specified; these may be authors, subject words or phrases, numbers, etc.

-----  
**COMMENTS:**

- Even if only one or two keywords are needed to do a search, a nonsense term or a repeat must be supplied to total three, for example:  

```
CALL FINDOR ('NO'USSR'ELITES'XYZ')END
-OR-
CALL FINDAND ('NO'USSR'ELITES'USSR)END
```
- If the reference does not report one of three NU access procedures (ISLIB, ICPRLTB, UPDATE), then the study is not in NU's holdings.
- If the user chooses not to list the reports at the terminal, hard copy can be printed by:  

```
$DISPOSE, DISPLAY, PR.
```
- When the search is completed, the user can make additional searches within the same search logic by merely repeating step III and altering the four parameters in parentheses as needed. To change search logic, begin at step II and continue on.

FIGURE 1: How to Find Studies in ICPSR and NU Holdings Using SEEKER Online

'ABORT' in study abstracts is given in Figure 2. SEEKER has also retrieved four studies, but only two are the same as those listed in the Guide subject index. Both SEEKER and the index identify the 1974 General Social Survey and a 1969 survey for CBS Reports as pertaining to "abortion," a word mentioned in the abstracts of both studies. But the subject index to the Guide also lists the 1972 American National Election Study and the cumulated 1972-78 General Social Survey as dealing with abortion, which SEEKER misses because the word did not appear in the abstracts. On the other hand, SEEKER retrieved a 1975 General Social Survey (study 7527) and a Dutch Election Study (7261) that mention abortion in their abstracts but were not listed in the subject index.

Results of this search (and others not reported here) indicate that a computer search of ICPSR study abstracts is a useful supplement to the subject index in the Guide to Resources and Services. It cannot replace the subject index, for the index cites some studies whose abstracts do not use the words that the SEEKER procedure searches. Based only on retrieval effectiveness, the index and SEEKER seem about equal. They may be evaluated differently on other grounds, such as ease of use and return for effort expended. Searching the printed subject index is familiar and straightforward for the inexperienced user, while SEEKER requires new learning. However, the computer search is faster in identifying studies, which are only cited by page and number in the index. Seated at a terminal using SEEKER, the user can rapidly identify the studies, print the abstracts, and also learn of their availability and access mode at Northwestern. Because of speed and convenience, SEEKER is preferable for knowledgeable users, especially when the number of studies listed in the subject index is large.

A computer search of the second file--VARLIST-- offers an even more powerful data retrieval capability, for the search is conducted at the data item level. Even if a study does not focus on abortion, it may contain questions probing attitudes toward abortion. Searching at the data item level within studies can locate such questions. RIQS has been used before to search survey questions by Caes Middendorf at the Steinmetz Archives, University of Amsterdam. This application, however, met with limited success, for many words used in phrasing questions were not useful in retrieving information. The VARLIST file, on the other hand, does not contain actual survey questions but rather ICPSR-supplied descriptions of variables formulated from the questions. Consider this line of questioning about abortion from the 1974 General Social Survey:

"Please tell me whether or not you think it should be possible for a pregnant women to obtain a legal abortion . . . If there is a strong chance of serious defect in the baby? . . . If she is married and does not want any more children?" (and so on)

These questions yielded six variables, described in the codebook for study 7341 as follows:

R'S OPINION: ABORTION (V131-136)

131 ABORTION AVAILABLE IF CHANCE OF SERIOUS DEFECT IN BABY

132 ABORTION AVAILABLE IF WOMAN WANTS NO MORE CHILDREN

Seeker.

RIGS ... NORTHWESTERN UNIVERSITY  
(VERSION 2.1 - 12/15/79)

FILE HAS THE NAME GUIDE  
AND CONTAINS 807 RECORDS -- 791 ACTIVE RECORDS  
CREATED JAN 25, 1981  
EDITION 7 CREATED MAR 12, 1981  
ARE YOU USING A PRINTING TERMINAL?  
? yes

RIGS MONITORS SEARCH TEXT. PLEASE ENTER YOUR NAME AND DEPARTMENT.  
? Ann Janda, Vogelback Computing Center

DO YOU WANT A DESCRIPTION OF YOUR FILE  
? no

ENTER SEARCH COMMAND OR TYPE HALT

? retrieve finder finder  
RETRIEVED PROCEDURE FINDER  
? call finder ('yes' 'abort' 'abortion' 'abortions')end

SEARCHING INITIATED

NO. OF REPORTS ON DISPLAY FILE = 4

DO YOU WANT THE DISPLAY REPORTS LISTED  
? yes

TITLE: ICPSR INSTRUCTIONAL SUBSET; GENERAL SOCIAL SURVEY, 1975.  
ICPSR NUMBER: 7527  
AUTHOR(S): NATIONAL OPINION RESEARCH CENTER  
CATEGORY: INSTRUCTIONAL PACKAGES AND COMPUTER PROGRAMS  
- INSTRUCTIONAL PACKAGES - ICPSR  
INSTRUCTIONAL SUBSETS

NO ACCESS LIBRARY:  
FILE NAME :  
STUDY ABSTRACT: THE NATIONAL DATA PROGRAM FOR THE SOCIAL SCIENCES IS A DATA DIFFUSION PROJECT AND A PROGRAM OF SOCIAL INDICATOR RESEARCH FUNDED BY THE NATIONAL SCIENCE FOUNDATION. THE PROGRAM WAS DESIGNED AND CARRIED OUT OVER A FIVE-YEAR PERIOD BY THE NATIONAL OPINION RESEARCH CENTER (NORC), UNIVERSITY OF CHICAGO. THE SPRING, 1975, GENERAL SOCIAL SURVEY, ADMINISTERED IN MARCH AND APRIL, 1975, WAS THE FOURTH IN A SERIES OF FIVE ANNUAL SURVEYS WHICH BEGAN IN 1972 AND CONCLUDED IN 1976. THE STUDY INCLUDED 1,490 RESPONDENTS AND 237 VARIABLES. THIS SUBSET CONTAINS SEVENTY-SIX VARIABLES. THE 1975 INTERVIEWS INCLUDED ITEMS SELECTED BY THE NORC STAFF AND AN ADVISORY PANEL OF SOCIOLOGISTS AS BEING 'MAINSTREAM' INTERESTS OF ACADEMIC SOCIOLOGY. ASIDE FROM STANDARD PERSONAL DATA ITEMS, IT COVERED SUCH AREAS OF INTEREST TO SOCIAL SCIENTISTS AS THE FAMILY, SOCIO-ECONOMIC STATUS, SOCIAL MOBILITY, SOCIAL CONTROL, RACE RELATIONS, SEX RELATIONS, AND MORALE. THIS ICPSR SUBSET BEGINS WITH SEVERAL ITEMS OF PERSONAL DATA, SUCH AS REGION AND SIZE OF PLACE WHERE THE RESPONDENT LIVES, INFORMATION ABOUT THE RESPONDENT'S FAMILY BACKGROUND, THE EMPLOYMENT STATUS OF THE RESPONDENT AND SPOUSE, SEX, RACE, RELIGION, AND POLITICAL PARTY IDENTIFICATION. FOLLOWING A FEW QUESTIONS REGARDING FAMILY FINANCES AND WORK SATISFACTION, RESPONDENTS WERE ASKED A NUMBER OF QUESTIONS ABOUT GENDER ROLES, INCLUDING OPINIONS ON WHETHER WOMEN 'SHOULD TAKE CARE OF RUNNING THEIR HOMES' AND ON THE EMOTIONAL SUITABILITY OF MEN AND WOMEN TO POLITICS. NEXT RESPONDENTS WERE ASKED A NUMBER OF QUESTIONS RELATED TO ABORTION AND THE CONDITIONS UNDER WHICH IT MIGHT BE VIEWED AS JUSTIFIED. RESPONDENTS WERE ALSO ASKED SEVERAL QUESTIONS TO GAUGE THEIR ATTITUDES TOWARD PORNOGRAPHY AND ITS CONTROL, TOWARD PRE-MARITAL SEXUAL RELATIONS, AND TOWARD PRE-MARITAL SEXUAL RELATIONS.

TITLE: DUTCH ELECTION STUDY, 1970-1973.  
ICPSR NUMBER: 7261  
AUTHOR(S): HEUNKS, FELIX \* JENNINGS, M. KENT \* MILLER, WARREN E. \* STOUTHARD, PHILIP C. \* THOMASSEN, JACQUES

CATEGORY: MASS POLITICAL BEHAVIOR AND ATTITUDES - HIST. AND CONTEMPORARY ELECTORAL PROCESSES - ELECTION STUDY SERIES - NATIONS OTHER THAN US  
ISLIB

NO ACCESS LIBRARY:  
FILE NAME : DES70 (1970 ONLY)  
STUDY ABSTRACT: THE ELECTION STUDY WAS CONDUCTED AS A THREE-WAVE PANEL WITH THE FIRST WAVE CONDUCTED IN MARCH-JULY, 1970 (1,838 RESPONDENTS), THE SECOND WAVE IN APRIL, 1971, AFTER THE PARLIAMENTARY ELECTION (1,262 OF THE FIRST WAVE RESPONDENTS), AND THE THIRD WAVE AFTER THE PARLIAMENTARY ELECTION OF NOVEMBER, 1972 (972 OF THE ORIGINAL RESPONDENTS). A QUESTIONNAIRE WAS MAILED TO RESPONDENTS WHO REFUSED THE THIRD INTERVIEW; A TOTAL OF 356 RESPONSES WERE OBTAINED FROM THIS MAILING. THERE ARE A TOTAL OF 1,838 RESPONDENTS AND 766 VARIABLES IN THIS STUDY. THE THREE-WAVE ELECTION STUDY FOCUSES ON THE PARTISAN ORIENTATIONS OF THE DUTCH PEOPLE. QUESTIONS RELATED TO PARTY IDENTIFICATION, ELECTORAL CHOICES IN THE THREE PARLIAMENTARY ELECTIONS AS WELL AS LOCAL ELECTIONS, CANDIDATES AWARENESS, ISSUE IMPORTANCE, AND POLITICAL ACTIVITIES WERE ASKED IN EACH WAVE. IN ADDITION, WAVE TWO CONCENTRATED ON THE CONCEPT OF REPRESENTATION: QUESTIONS ABOUT CITIZENS' PERCEPTION OF THE STANCE OF POLITICAL PARTIES WITH RESPECT TO SUCH NATIONAL ISSUES AS ABORTION, CIVIL DISTURBANCES, AID TO DEVELOPING NATIONS, INCOME DISTRIBUTION, TAXATION AND DEFENSE SPENDING, AS WELL AS OPINIONS ON THE NECESSARINESS OF REPRESENTATION TO CITIZEN DEMANDS.

TITLE: GENERAL SOCIAL SURVEY, 1974.  
ICPSR NUMBER: 7341  
AUTHOR(S): NATIONAL OPINION RESEARCH CENTER  
CATEGORY: SOCIAL INDICATORS - US

NO ACCESS LIBRARY: ISLIB  
FILE NAME : GENSOCC3  
STUDY ABSTRACT: AS IN 1972 AND 1973, THE 1974 INTERVIEW INCLUDES ITEMS SELECTED BY THE NORC STAFF AND AN ADVISORY PANEL OF SOCIOLOGISTS AS 'MAINSTREAM' INTERESTS OF ACADEMIC SOCIOLOGY. IN ADDITION TO STANDARD PERSONAL DATA ITEMS, THE 1974 SURVEY COVERS SUCH AREAS OF INTEREST AS THE FAMILY, SOCIO-ECONOMIC STATUS, SOCIAL MOBILITY AND MORALE. ABOUT TWO-THIRDS OF THE QUESTIONNAIRE PROBED FOR ATTITUDES AND OPINIONS CONCERNING QUALITIES OF A JOB, SATISFACTION WITH LIFE, ROLES OF WOMEN, BIRTH CONTROL AND ABORTION, SEX RELATIONS, RACE RELATIONS, SOCIAL CONTROL ISSUES, AND CIVIL LIBERTIES. INFORMATION WAS OBTAINED FROM 1484 RESPONDENTS IN THE FOLLOWING AREAS:

TITLE: CBS REPORTS: GENERATIONS APART.  
ICPSR NUMBER: 7345  
AUTHOR(S): COLUMBIA BROADCASTING SYSTEM  
CATEGORY: SOCIAL INSTITUTIONS AND BEHAVIOR - SOCIALIZATION STUDENTS AND YOUTH - US

NO ACCESS LIBRARY:  
FILE NAME :  
STUDY ABSTRACT: 1366 COLLEGE STUDENTS AND NON-COLLEGE YOUTH BETWEEN 17 AND 23, 5 CARDS OF DATA PER RESPONDENT AND 288 VARIABLES. THIS STUDY WAS CONDUCTED IN 1969 BY DANIEL YANKELOVICH, INC., FOR THE COLUMBIA BROADCASTING SYSTEM. THE RESULTS WERE BROADCAST MAY 20, 27 AND JUNE 3, 1969 IN THREE SECTIONS: 'QUESTION OF VALUES,' 'A PROFILE OF DISSENT,' AND 'THE YOUTH INTERNATIONAL.' A STUDY OF THE GENERATION GAP, THIS SURVEY CONTAINS QUESTIONS ON THE TYPE OF SOCIAL CHANGE AND SOCIETAL RESTRAINTS THE RESPONDENTS WOULD WELCOME OR REJECT. IN ADDITION THE RESPONDENTS WERE PROBED ON THEIR VIEWS OF THEIR PARENTS' VALUES AS WELL AS THEIR OWN. THEY WERE ASKED WHICH POLITICAL EVENTS HAD AFFECTED THEIR LIFE AND VALUES. ISSUES INQUIRED INTO INCLUDE ABORTIONS, SEXUAL RELATIONS, CIVIL DISOBEDIENCE, CRITICISM OF AMERICAN SOCIETY, CAREER GOALS, THE DRAFT, AND SEXUAL RELATIONS.

FIGURE 2: Output of SEEKER Search for Data on Abortion

- 133 ABORTION AVAILABLE IF WOMAN'S OWN HEALTH IS IN DANGER
- 134 ABORTION AVAILABLE IF POOR FAMILY CANNOT AFFORD CHILD
- 135 ABORTION AVAILABLE IF WOMAN IS RAPE VICTIM
- 136 ABORTION AVAILABLE IF WOMAN IS UNMARRIED

This more parsimonious and relevant text should be much better suited to keyword search and retrieval than the text of actual questions. The variable descriptions were entered directly into RIQS with a minimum of editing. Only the variables' range (V131-136) was added to the heading, and each variable was separated by an asterisk so that RIQS would treat each as a separate subitem. Of course, a good deal of codebook text (e.g., response

categories) was deleted also. Even this manual intervention was handled "automatically" with the Consortium codebook in a text editor. Because the VARLIST file is still under development, only 12 machine-readable codebooks have been converted for input to RIQS. Although only survey codebooks were selected for initial conversion, nothing would prevent including codebooks from aggregate data studies. To search at the data item level, a different capability of RIQS was used. We now wanted not only to locate the studies that contained questions about abortion but also to identify the specific variables in the studies that contained the information. For this, we used the familiar keyword-in-context technique (KWIC), an optional procedure in RIQS, which and is offered in other information retrieval systems. The KWIC procedure selected all the subitems containing the keywords 'ABORTION,' 'ABORTIONS,' and 'ABORT,' and printed the subitems alphabetized by keyword within context. This is also an online search procedure, but because the data set is still under development, it has not yet been "simplified" for the inexperienced user. The results for our sample search are reproduced in Figure 3.

	<u>Keyword</u>	<u>Study</u>
	↓	↓
S R HEARD SUPREME COURT DECISION ABOUT	ABORTION	7315
450 R'S OPINION ON	ABORTION	7655
D	162 ABORTION ALLOWABLE IF CANNOT AFFORD CHIL	7315
	159 ABORTION ALLOWABLE IF DEFECT IN BABY	7315
REN	160 ABORTION ALLOWABLE IF WANT NO MORE CHILD	7315
IM	163 ABORTION ALLOWABLE IF WOMAN IS RAPE VICT	7315
	164 ABORTION ALLOWABLE IF WOMAN IS UNMARRIED	7315
DANGER	161 ABORTION ALLOWABLE IF WOMAN'S HEALTH IN	7315
801, 1976)	ABORTION AND SEX DISCRIMINATION (V3796-3	7607
DEFECT IN BABY	131 ABORTION AVAILABLE IF CHANCE OF SERIOUS	7341
DEFECT IN BABY	126 ABORTION AVAILABLE IF CHANCE OF SERIOUS	7367
DEFECT IN BABY	116 ABORTION AVAILABLE IF CHANCE OF SERIOUS	7398
DEFECT IN BABY	126 ABORTION AVAILABLE IF CHANCE OF SERIOUS	7573
AFFORD CHILD	134 ABORTION AVAILABLE IF POOR FAMILY CANNOT	7341
AFFORD CHILD	129 ABORTION AVAILABLE IF POOR FAMILY CANNOT	7367
AFFORD CHILD	119 ABORTION AVAILABLE IF POOR FAMILY CANNOT	7398
AFFORD CHILD	129 ABORTION AVAILABLE IF POOR FAMILY CANNOT	7573
IM	135 ABORTION AVAILABLE IF WOMAN IS RAPE VICT	7341
IM	130 ABORTION AVAILABLE IF WOMAN IS RAPE VICT	7367
IM	120 ABORTION AVAILABLE IF WOMAN IS RAPE VICT	7398
IM	130 ABORTION AVAILABLE IF WOMAN IS RAPE VICT	7573
	136 ABORTION AVAILABLE IF WOMAN IS UNMARRIED	7341
	131 ABORTION AVAILABLE IF WOMAN IS UNMARRIED	7367
	121 ABORTION AVAILABLE IF WOMAN IS UNMARRIED	7398
	131 ABORTION AVAILABLE IF WOMAN IS UNMARRIED	7573
E CHILDREN	132 ABORTION AVAILABLE IF WOMAN WANTS NO MOR	7341
E CHILDREN	127 ABORTION AVAILABLE IF WOMAN WANTS NO MOR	7367
E CHILDREN	117 ABORTION AVAILABLE IF WOMAN WANTS NO MOR	7398
E CHILDREN	127 ABORTION AVAILABLE IF WOMAN WANTS NO MOR	7573
IS IN DANGER	133 ABORTION AVAILABLE IF WOMAN'S OWN HEALTH	7341
IS IN DANGER	128 ABORTION AVAILABLE IF WOMAN'S OWN HEALTH	7367
IS IN DANGER	118 ABORTION AVAILABLE IF WOMAN'S OWN HEALTH	7398
IS IN DANGER	128 ABORTION AVAILABLE IF WOMAN'S OWN HEALTH	7573
	3796 WHEN SHOULD	7607
PARTY WOULD SUPPORT AMENDMENT TO MAKE	ABORTION BE ALLOWED	7607
ORD THE CHILD	ABORTION ILLEGAL	7607
EFFECT IN BABY	59 ABORTION SHOULD BE POSSIBLE IF CAN'T AFF	7309
ORE CHILDREN	56 ABORTION SHOULD BE POSSIBLE IF SERIOUS D	7309
RAPE VICTIM	57 ABORTION SHOULD BE POSSIBLE IF WANT NO M	7309
UNMARRIED	60 ABORTION SHOULD BE POSSIBLE IF WOMAN IS	7309
HEALTH DANGER	61 ABORTION SHOULD BE POSSIBLE IF WOMAN IS	7309
	58 ABORTION SHOULD BE POSSIBLE IF WOMAN'S H	7309
	ABORTION (V116-121)	7398
	ABORTION (V126-131)	7367
	ABORTION (V126-131)	7573
	R'S OPINION: ABORTION (V131-136)	7341
	R'S OPINION- ABORTION (V158-164)	7315
	R'S OPINION- ABORTION (V56-62)	7309
AMENDMENT TO THE CONSTITUTION TO ILLEGALIZE	ABORTIONS	7660
AMENDMENT TO THE CONSTITUTION TO MAKE	ABORTIONS ILLEGAL	7660

FIGURE 3: KWIC Output of RIQS Search of Data Items in the VARLIST File

Whereas the subject index to the Guide and the SEEKER search together identified six different studies that contained attitudinal data toward abortion, our search of only twelve studies at the data item level in VARLIST found nine studies and only one was duplicated in the other six. No doubt this search would have found the other five studies as well if their codebooks had been among those converted for VARLIST. One wonders how many studies we would have found if more codebooks had been converted. This search with RIQS at the data item level suggests that users must be missing a great deal of relevant data by relying on descriptions of data files instead of looking at data items. This lapse in information would occur whether the search is conducted manually, using a subject index, or with computer assistance.

#### Summary and Conclusion

This study discussed the system in use at Northwestern University for locating data in the holdings of the Inter-University Consortium for Political and Social Research. The system uses a general purpose information retrieval program, RIQS, which operates only on CDC computers. Its principles, however, are easily adaptable for implementation with other retrieval programs at member institutions of the Consortium. The key to the system is the use of Consortium tapes containing the textual material for searching. The machine-readable material from the ICPSR is of two forms: (1) the ICPSR Guide to Resources and Services and (2) machine-readable codebooks as distributed with ICPSR data sets.

Two approaches to searching ICPSR holdings were discussed. A simple, user-oriented routine called SEEKER searched at the data file level, using the study abstracts with reference citations as reported in the Guide. SEEKER was applied to a hypothetical search for a student interested in data on attitudes toward abortion. This technique of searching for keywords (i.e., 'abortion') in study abstracts missed some studies that were listed in the subject index to the Guide, but it also found some that were not indexed. The other approach, searching at the data item level, was applied to the variable descriptions in a small number of Consortium codebooks converted for RIQS processing. This retrieval technique disclosed many more studies containing data on abortion than were found with both the manual and the computer methods operating at the data file level.

The paper began by noting that every solution seems to introduce a new problem. Using the computer to search a data base of ICPSR study descriptions was found to be a useful supplement to the subject index in the Guide and, from the standpoint of speed and convenience, even preferable. But using the computer this way immediately suggested its application to searching the variable descriptions themselves. A new problem arose from learning that this data base and technique constitute a more powerful search strategy that deserves the additional expenditure of time and resources to develop.

#### REFERENCES

- BISCO, Ralph L. (1967)  
"Social Science Data Archives: Progress and Prospects," Social Science Information, 6 (February), 39-74.
- CONGER, Linda D. (1976)  
"Data Reference Work with Machine Readable Data Files in the Social Sciences," Journal of Academic Librarianship, 2 (May) 60-65.
- DEUTSCH, Karl W. (1967)  
"The Impact of Complex Data bases on the Social Sciences," in Ralph L. Bisco (ed.) Data Bases, Computers, and the Social Sciences. New York: Wiley, 1970. Pp.19-41.
- DODD, Sue A. (1977)  
"Cataloging Machine-Readable Data Files -- A First Step?" Drexel Library Quarterly, 13 (January), 48-69.
- DODD, Sue A. (1978)  
"Building a Bibliographic Data Base for Machine-Readable Data Files -- with Related Services and Products." Paper presented at the Annual IASSIST Conference, February 8-11, Itasca, Illinois.
- FERGUSON, Douglas (1977)  
"Social Science Data Files, the Research Library and the Computing Center," Drexel Library Quarterly, 13 (January), 70-81.
- HOFFERBERT, Richard I. and Jerome M. CLUBB (1976)  
"Introduction," American Behavioral Scientist, 19 (March/April), 381-386.
- JANDA, Kenneth (1980)  
"Cross-National Data for Social Research: 50 Studies in the ICPSR Archives." Paper prepared for delivery at the U.S./Soviet Symposium on Social Science Information, June 4-7, Ann Arbor, Michigan.
- MITTMAN, Benjamin and Lorraine Borman (eds.) (1975)  
Personalized Data Base Systems. Los Angeles: Melville.
- MOCHMANN, Ekkehard (1974-1975)  
"Information Access at the Data Item Level: Approaches to Indicator Retrieval from Survey Archive Data Bases," SIGSOC Bulletin, 6 (Fall/Winter), 47-55.

NIELSEN, Per (1974-1975)  
"Information Access at the Data File  
Level: Documentation Prerequisites on the  
File-Level Data Base Inquiry Process,"  
SIGSOC Bulltin, 6 (Fall/Winter), 39-46.

PETERS, Paul (1974-1975)  
"Describing a Social Science Data  
Information System, Networks and  
Components," SIGSOC Bulletin, 6  
(Fall/Winter), 6-25.

ROKKAN, Stein (1966)  
Data Archives for the Social Sciences.  
Paris: Mouton.

SIDEL, Philip S. (1980)  
"Social Science Data Archives,"  
Encyclopedia of Library and Information  
Sciences, 28, pp.45-60.

SOBAL, Jeff (1981)  
"Teaching with Secondary Data," Teaching  
Sociology, 8 (January), 149-170.

TANNENBAUM, Eric (1979)  
"Local Data Libraries for Quantitative  
Political Analysis," Teaching Political  
Science, 6 (July), 413-438.

WHITE, Howard D. (ed.) (1977)  
Reader in Machine-Readable Social Data.  
Englewood, Colorado.