## In this issue:

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Editor's Log</td>
<td>1</td>
</tr>
<tr>
<td>Data Archives and Libraries: New Challenges</td>
<td>3</td>
</tr>
<tr>
<td>Jacqueline McGee</td>
<td></td>
</tr>
<tr>
<td>Organizing a Data Library</td>
<td>7</td>
</tr>
<tr>
<td>Ann E. Gerken</td>
<td></td>
</tr>
<tr>
<td>The Care and Feeding of Magnetic Tapes</td>
<td>11</td>
</tr>
<tr>
<td>Patricia A. Reslock</td>
<td></td>
</tr>
<tr>
<td>Data Service in a Computer Center</td>
<td>14</td>
</tr>
<tr>
<td>Gertrude J. Lewis</td>
<td></td>
</tr>
<tr>
<td>Utilities</td>
<td>18</td>
</tr>
<tr>
<td>Jacqueline McGee</td>
<td></td>
</tr>
<tr>
<td>Aspects of Data Management</td>
<td>23</td>
</tr>
<tr>
<td>Ilona Einowski</td>
<td></td>
</tr>
<tr>
<td>Data Service in a Library Setting</td>
<td>26</td>
</tr>
<tr>
<td>Bliss Beckman Siman</td>
<td></td>
</tr>
</tbody>
</table>
In this issue:

EDITOR'S LOG ..................... 1
DATA ARCHIVES AND LIBRARIES: NEW CHALLENGES
    Jacqueline McGee ................ 3
ORGANIZING A DATA LIBRARY
    Ann E. Gerken ................... 7
THE CARE AND FEEDING OF MAGNETIC TAPES
    Patricia A. Reslock ............. 11
DATA SERVICE IN A COMPUTER CENTER
    GERTRUDE J. LEWIS ............... 14
UTILITIES
    Jacqueline McGee ................. 18
ASPECTS OF DATA MANAGEMENT
    Ilona Einowski ................... 23
DATA SERVICE IN A LIBRARY SETTING
    Bliss Beckman Siman .............. 26
EDITORIAL INFORMATION

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

INFORMATION FOR AUTHORS

The QUARTERLY is published four times yearly. Articles and other information should be typewritten and double-spaced. Each page of the manuscript should be numbered. The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago Press Manual of Style or Kate L. Turabian's Manual for Writers. If the contribution is an announcement of a conference, training session, or the like, the text should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event. Book notices and reviews should not exceed two double-spaced pages. Deadlines for submitting articles are six weeks before publication. Manuscripts should be sent in duplicate to the Editor:

Elizabeth Stephenson
Institute for Social Science Research
University of California
405 Hilgard Avenue
Los Angeles, California 90024
USA
(213) 825-0716 or (213) 825-0711

Book reviews should be submitted in duplicate to the Book Review Editor:

Kathleen M. Heim
School of Library and Information Science
Louisiana State University
Coates Hall, Room 267
Baton Rouge, Louisiana 70803
USA
(504) 388-3158

Dates of publication:
Volume 8 No. 1 - January 1984
Volume 8 No. 2 - April 1984
Volume 8 No. 3 - September 1984
Volume 8 No. 4 - November 1984

Key Title: Newsletter - International Association for Social Science Information Service and Technology

ISSN - United States: 0739-1137

Copyright © 1982 by IASSIST. All rights reserved.
Public data like census data, administrative micro-data and survey data, offer many opportunities for research in the social sciences and related disciplines. In the past five years new developments in computer software and hardware have accelerated the technical availability of data material. Computernetworks, 'intelligent' terminals, micro-computers, on-line databases, sophisticated software packages etc. make it technically possible to access an enormous amount of data for scientific research.

With these developments new problems emerged concerning the access of the data. Existing regulations for the control of the flow of data proved to be inadequate. Researchers tried to find new ways for optimal usage of the technical possibilities to get access to public data.

The next IFDO/IASSIST conference in Amsterdam will be an excellent opportunity for people who are dealing with public data to present new developments in this area, and to discuss the related problems. The meetings will include the discussion of papers on a variety of topics of interest to social scientists, data archivists, librarians, research administrators, government records managers and users of data banks.

CONFERENCE FORMAT

The conference will include plenary sessions and concurrent sessions with presentations and demonstrations. One day of the conference will be devoted to workshops on specific topics like census software packages, statistical programs for micro-computers and international data. Conference language: English.

The registration fee will be Dfl. 300 (app. US$ 100). The fee includes conference activities, workshops, a book containing the outlines of the presented papers, coffee- and tea-breaks, a reception and a number of meals. The meetings are planned in the conference rooms of the Grand Hotel Krasnapolsky, in the very center of Amsterdam.

CALL FOR PAPERS

Papers are being solicited on the various aspects of the theme of Public access to public data as described above. Abstracts of papers for the conference should be submitted to the conference program committee before December 1, 1984. Abstracts should be typed in English, with a maximum of 500 words.

CONFERENCE ORGANIZATION

The International Federation of Data Organizations for the Social Sciences (IFDO) and the International Association for Social Science Information Service and Technology (IASSIST) cooperate in the organization of the Amsterdam conference. The conference will be hosted by the Steinmetz Archives (the Dutch data archive), which is a department of the Social Science Information and Documentation Center.

For additional information write to:

STEINMETZ ARCHIVES
IFDO/IASSIST CONFERENCE
Herengracht 410-412
1017 BX Amsterdam
THE NETHERLANDS
Telephone (20) 225061

Please indicate whether you plan to submit an abstract!
Since 1984 marked the 10th anniversary of IASSIST as an organization, I felt it would be useful to focus an issue of the QUARTERLY on the present concerns of social science data archives. In the articles that follow, Anne Gerken, Bliss Siman, and Gert Lewis discuss the organization of an archive in different settings. Patricia Reslock and Jacqueline McGee have prepared information on utilities and techniques for handling machine-readable information. An overall picture of data management and the changing role of archives is described in the articles by Ilona Einowski and Jacqueline McGee.

These articles were prepared from papers presented and workshops conducted at the last two IASSIST annual conferences. The sessions focused on the management of archives and data, with representatives from a variety of settings who reported on the functions and organization of their particular facilities. I hope that the articles can give readers at least the flavor of the lively discussion and interaction that was generated in these gatherings.

Finally, many thanks go to Ilona Einowski and Bliss Siman who were the co-editors for this issue of the QUARTERLY. They did much of the tedious and time-consuming work of gathering and preparing the articles. It is due to their efforts that we were able to bring together the experiences of other archivists and present them here.

Best wishes,

Elizabeth Stephenson

Elizabeth Stephenson
The comments presented by me today will necessarily reflect my view of the research community from my position in the Rand Computation Center Data Facility. It is hoped some of my remarks will also reflect a more general situation.

The topics to be discussed in this paper are:

1. The impact of government policy on the establishment and maintenance of data archives;
2. The changing role of data archives; and
3. The changing requirements of research clientele.

If my remarks are to be pertinent for this session I feel it is necessary to provide for you a brief historical sketch of the Rand Corporation. Not only to clarify my position on this subject, but to reflect on the past, in order to describe more clearly the present, and to contemplate the future.

Rand History

The Rand Corporation was organized in 1946 as the RAND Project when the then Army Air Forces (AAF) awarded to the Douglas Aircraft Company, a contract whose purpose was to provide scientific advice and recommendations to the AAF. Two years later, to prevent what was seen by some as a possible conflict of interest, the RAND Project became the Rand Corporation with private funds from the Ford Foundation. At first, work at Rand concentrated mainly on projects of national security issues. However, some research did address domestic policy issues especially in areas of transportation, water supply, mental health and local government needs.

Later, in 1969, at the request of Major Lindsay of New York City, the New York Rand Institute was formally established. The work of the Institute proposed to aid the city in identifying opportunities and problems and assist with the decision-making process. The Institute work included the active involvement of the city officials. It was here that many of the domestic research projects later undertaken at Rand had their beginnings.

Today the Rand Corporation is a private nonprofit research institution. Also housed at Rand is the Rand Graduate Institute. The Institute offers a graduate program leading to a doctoral degree in policy analysis. Students participate in formal academic study and also receive on-the-job training in applied research while participating in Rand projects. Research projects are conducted in three major divisions—Project Air Force, National Security, and Domestic Research—and the Civil Justice Institute. The first of these divisions, Project Air Force, is still in operation after 37 years. The largest
single activity at Rand, Project Air Force contributed approximately 37 percent of the total revenue in Fiscal Year 1983. The second division, National Security, contributed 28 percent of the fiscal year revenue. The third division, Domestic Research, contributed 29 percent and the Civil Justice Institute contributed 5 percent.

Research staff for all divisions at Rand are housed in six research departments, Behavioral Science, Economics, Engineering and Applied Science, Information Science, Political Science and System Science. Supporting departments include research Libraries, Publications and Computer Services.

Research projects draw personnel from all of the six departments. Policy analysis and research at Rand has been described as "interdisciplinary." Interdisciplinary research requires that members of the research projects interact and integrate their separate disciplines.

Data Facility

The Data Facility was initiated at Rand about ten years ago. Originally the Facility was envisioned as a much larger and more costly operation which would have included data management for projects, but a decision was made not to invest too heavily in the management of data and an attempt to identify possible data files for archiving became the Facility's primary focus at that time.

The Data Facility (machine readable data file--MRDF--archive) resides in the Computer Service Department. The Facility staff assists projects in identifying and acquiring data files from both Rand and non-Rand sources, and serves as a central clearing house for the acquisition, dissemination, control and storage for MRDF. The Data Facility staff attempts to acquire and archive data files with a high probability of future use or files which are considered generally useful for a wide variety of research applications.

The Data Facility staff assists researchers with proposals by supplying information about the availability of data, in the archive or elsewhere, the location or existence of variables or data and statistical reference sources. Project programmers and researchers using archived data frequently request assistance with the use of a particular data file, clarification of codes, or additional reference sources. Our facility also maintains separate storage for software products, Rand-generated programs and models, user manuals and references.

During the Facility's developing years the large government sponsored social experiments and demonstration projects were underway at Rand. The Housing Assistance Supply Experiment, the Health Insurance Study, and the National Preventive Dentistry Demonstration Project were a few of the large studies undertaken at Rand and in some cases are only now coming to a close. I suspect that many other similar large data collection projects elsewhere led to the establishment of some of today's archives.

Government Influence

To most of us the impact of government policies on the establishment and maintenance of data archives is fairly obvious. A careful inspection of many archive inventories would indicate that much of the data collected and utilized in the research process begins with a government agency. In fact, it would be difficult to locate many data files that were created without at least some government funds.

As an example, the Rand Data Facility maintains a collection of census data administrative records from the Bureau of Labor statistics and Social Security
Administration, survey data from the Department of Health and Human Resources, the National Center for Education Statistics, the Economic Census from the Bureau of Census and many other surveys and administrative data files. The most frequently researched stenographic data bases in the Data Facility are the National Longitudinal Labor Market Experience surveys, the Panel Study of Income Dynamics and the two student data files, the High School Class of 1982 and the High School and Beyond. All of these projects are supported with government funds. In recent years government agencies funding research have required the creation of public use files to be produced and archived and instituted at the completion of research projects.

Lately research centers in the United States have all felt the effects of the federal budget cuts not only in the granting of funds for research but in the collection and dissemination of data by federal agencies. Many agencies are reducing the number of surveys they are conducting, reducing the number of data files they distribute, or reducing staff and equipment to the point where service to the user community is affecting the acquisition process.

Perhaps because of these budget cuts, our archive has experienced an increase in the demand for existing data for use in secondary analysis. Certainly, there has been an increase in the number of queries about the holdings of our archive by researchers anticipating to lower the cost of a project by finding data to suit his needs.

Archive Future Outlook

Funding Research. It would be safe to assume that the process of funding research is changing and is a trend that will probably continue for the foreseeable future.

Some changes in research funding have taken place at Rand in recent years. For instance, the Rand Civil Justice Institute was established in 1979 to perform independent, objective policy analysis on the American Civil Justice system. The purpose of the Institute is to help make the civil justice system more efficient and more equitable by supplying policymakers with the results of empirically based, analytic research. An example of the projects undertaken by the Institute include occupational disease, determinants of delay to case disposition, history of court congestion and delay, selection disputes for litigation, private costs of civil cases, and inflation and jury awards.

The Institute is funded by a broad base of contributors in the private sector. Since its inception over 200 organizations have made contributions to the Institute, including property-casualty insurance companies, other major corporations, and trade and professional associations.

In addition, in 1982 the Private Sector Sponsors Program was initiated at Rand to facilitate private sector support of research benefitting both industry and government. Research and related activities in this new program will be conducted within the Domestic Research Division under regular Rand procedures and organizational structures.

When the Data Facility was first established, very little Rand Project data was archived. The archive holdings were generally data files acquired from non-Rand sources. Today we are gradually acquiring Rand-generated data from completed projects and assisting a much larger percentage of projects with their acquisition of data. Proposal support remains active.

If privately funded research is a trend, how will this affect the data archive? In the case of the research in the civil justice area, machine-readable data rarely exists. It is my hope that the ongoing projects will
create data in machine-readable form to be archived for future use.

Proliferation of Information. Recently a researcher described for me how five years ago he could search any library and find everything he needed to know about a particular subject. Today, the researcher added, the information available is more specific and specialized and there are far too many sources for him to investigate all of them.

I hope our archive can assist Rand researchers with diffusing and filtering and in some cases evaluating information that may be of interest to them. It is my hope that the Data Facility provides our clients with information on occasion which they might not have had or perhaps would not have had in as timely a manner.

Changing Technology. It appears the trend toward micro-computer use is going to continue. Not only will it continue, but rapidly expand. There are many different opinions as to the effect this will have on the research process and the use of mainframe computers.

One theory is that as the micro use goes up so will the use of the mainframe. I believe at this point there appears to be some justification for this theory. At this stage in technology it is still necessary for any but the smallest data sets to be downloaded from a mainframe.

However, I have had requests for information about data on floppy disks and also researchers who had not previously been so inclined now are requesting information about software packages and utilities on the mainframe, in order to "talk" to it from their micro.

Conclusions

Research is often conducted in a departmentalized manner. This is probably more true in an academic setting than at Rand, but to some extent, it is also true here. Researchers frequently depend on colleagues for information about data. Research conducted in a departmentalized manner may cause some information not to be as readily available. It is here the data archive can best assist in a research center.

The data archive is a storage of information about data files, data sources, and data information from many different subject areas. Another valuable service an archive can provide is the institutional "memory," providing researchers with historical information about past data uses as well as sources of information about new data files.

Just as important as the hard copy documents is the individual who provides a human trail that allows researchers to follow new leads to people who have had experience with a particular set of data, or special knowledge about some of the files.

Because I need the contacts, and need to trace the human trail for the researcher, it is necessary to maintain contacts with other managers of machine-readable information. For this reason I value my membership in organizations such as IASSIST since these associations help me to maintain a standard of professional practice and expertise I might not otherwise obtain.
Historical Background and Development

The Cornell Institute for Social and Economic Research (CISER) is an interdisciplinary organization of Cornell faculty which seeks to support, strengthen and enrich the social and economic research community. In May 1981, CISER was founded to develop and support research programs and provide services and facilities required for research projects.

The CISER Data Archive was established in February 1982 in cooperation with the Cornell University Libraries to provide central access and management for social science data to researchers on and off campus. Data Archive staff provide professional information services, technical consultation, and research support. CISER also sponsors workshops and seminar series, peer review of research proposals, grant management, computing facilities, newsletters, and a directory of research interests of faculty in the social sciences at Cornell.

The CISER Data Archive was established upon the recommendations of a committee made up of representatives from four colleges at Cornell, the university libraries, Cornell Computer Services, and the New York State Cooperative Extension Service. The committee based its recommendations upon a survey of ten data archives located within research centers. Information was gathered on staff, collections, funding, space, and computing consulting.

The Data Archive's goals are to:

1. establish and maintain a centralized archive of machine-readable tapes and documentation;
2. acquire data and supporting documentation, coordinate buying consortia, fill gaps in data file holdings, and assure the safekeeping of archival holdings;
3. provide an information center with professional reference and computer consulting in social science data, defining information needs and providing research services; and
4. support the research and service missions of the institute and the university.

With the assistance of Cornell's Social Science Librarian, CISER opened the Data Archive in early February, 1982. A professional archivist joined the CISER staff and assumed administrative duties as well as the planning responsibilities for the development of the archive. Data files were acquired, policies, mailing lists and ordering procedures were established, and a survey of faculty was taken to identify data files on campus and those needed. The survey was helpful in locating data files to incorporate into the archive, in establishing contacts with researchers, and in developing a collection policy.
Since its opening in 1982, the Data Archive has grown significantly. Staff, services, and the holdings of machine-readable data files have expanded. The archive is an increasingly important research support facility on campus, providing essential services to a wide range of users. In addition to walk-in information and consulting services, the archive offers workshops, seminars, and classroom lectures on data file contents and use. The integration of computer consulting with data file reference service makes the archive a unique resource at Cornell. The staff is dedicated to the provision of continuous support to the educational and research activities in the social sciences, from data information to advanced analytical support.

Funding

CISER and the Data Archive are supported by allocated funds from five colleges at Cornell. The acquisitions portion of the budget is relatively small since most data sets are acquired through Cornell's membership in the Inter-university Consortium for Political and Social Research, through New York State Data Center affiliate status, and through other cooperative agreements with state and federal agencies. In addition, a number of files are donated to the archive by the faculty. A collection policy regarding acceptance of donations and purchase decisions is vital to the development of the archival holdings. Computing costs and tape storage costs are separately allocated.

Staffing

The staff of the archive consists of a professional data archivist, two full-time computer consultants, and a half-time data manager. Part-time student assistants are also available during the academic year.

The data archivist's duties include administration of the archive, data file evaluation and acquisition, data file information services, policy making, and the development of new services for social science researchers. The archivist works closely with the Cornell University Libraries to develop integrated information services, and communicates with the Computing Services staff in regard to technical developments and services. The archivist is a professional information specialist with an academic research library background.

The computer consultants provide statistical computing consultation and are responsible for the technical development of the data archive. They also work on contract for special data projects producing custom files, reports, analyses, and data management. The consultants have social science backgrounds with experience in statistical analysis and computerized research techniques.

The data manager is a half-time employee with tape management responsibilities. This person keeps inventories of tape contents and data users and oversees the addition and copying of tapes in the collection. Fundamental knowledge of the computer and tape management systems is required.

Student assistants perform tape management tasks and edit inventory files. Valuable management and secretarial assistance is also available.

Physical Environment

The archive is housed with the other offices of CISER. One large office houses the archivist, the data manager, and the library of technical documentation and reference materials. Other offices house the consultants. Each office has space for archive users to examine materials. Print materials can be taken out overnight. The consultants' offices have enough space for small group instruction and storage space for printouts and other records.
Equipment

As for local equipment, the archive has an IBM-PCl, will be acquiring additional microcomputers, and has two terminals which are used to communicate with the Cornell mainframe computers. The microcomputers are also used as terminals for communication and data transfer, as text processors, and for social science workstation development, including database management, graphics, and custom programming. Data on floppy diskettes are being distributed.

Public computing facilities in the building offer state-of-the-art graphics equipment, high speed printers, consultants, and technical manuals. CISER also has a computing facility in the building, with terminals for use by CISER members and their research assistants. Extensive microcomputer facilities are also located in the same building as the Data Archive including a software library and a demonstration facility.

Other hardware access includes an IBM 3081D, and a DEC 1020. The holdings of the archive are stored at the mainframe facility. Tapes are used on the IBM and also are exported for use on other systems accessible to Cornell researchers.

Sources of Data

The CISER Data Archive holds machine-readable data in the areas of demography, vital statistics, health, social surveys, labor and employment, occupation international trade, business, service industries, education, agriculture, and life studies and aging. The archive comprehensively collects New York State data.

CISER is an Affiliate of the New York State Data Center, and acquires many of its files from that source. Cornell is also a member of the Inter-univeristy Consortium for Political and Social Research (ICPSR) which provides the majority of non-Census files to Cornell. Longitudinal data are acquired from the Bureau of Economic Analysis. The archive receives data from numerous government agencies, both federal and state, and also acquires data files from other research institutions and survey centers. The contribution of research data files by Cornell faculty members have been central to the development of the data archive. The archive seeks continuous data deposits to build longitudinal strengths. Detailed collection development policies are developed in collaboration with faculty, librarians, and members of the Institute.

Dissemination of Information

To make users of the archive aware of the holdings, a title list of data files in the archive is updated and distributed bimonthly. Information about the archive, its holdings and services, is included in the bimonthly newsletter from CISER, called the syntheCISER. The archivist meets with faculty, graduate students, and staff to advertise and promote use of the archive, teach methods of identifying data files, and establish a network of data users. Future developments will include online directories of holdings and variable-level indexing of statistical data files. Cooperative cataloging of machine-readable records is being investigated. In addition, a model relationship with one of the Cornell University Libraries has been established whereby professional staff development, acquisitions, and information dissemination is coordinated.

Data are disseminated through tape and disk access on the IBM mainframe, through file transfer to tapes and diskettes, and in special data files and printouts provided on custom bases.
Users of the CISER Data Archive

The users of the archive represent the many colleges and departments at Cornell, and numerous off-campus organizations. Services are available to faculty, graduate and undergraduate students, staff, off-campus service agencies, private corporations, government agencies, and the greater Ithaca community. Access to the holdings of the archive on the mainframe are limited to those with Cornell computer accounts. Information and data delivery services are available to others on a contractual basis, except when data are restricted to use by the Cornell community. Fee structures have been developed to recover costs of some tasks. The growth of the archive is evident in the increasing numbers of walk-in and returning users.

Future Developments

There have been a number of developments that present challenges to CISER and affect the long-term plans for the archive and the Institute. Among the goals is the expansion of the archive collection and services. The archive staff's expertise in census and other federal data products has brought an increasing number of requests for special workshops, tabulations, and data file extracts of federal data files. As an Affiliate of the New York State Data Center, the archive frequently provides assistance to people throughout New York State. Requests for assistance with federal data from the Cornell community and off-campus institutions and organizations are expected to increase.

Researchers require subfiles and increased consultation for the larger longitudinal data sets and the microdata files in the data archive. An enlarged subsetting service and increased expertise in the management of heavily used data sets are being developed. In addition, technical support agreements, workshops, and classroom lectures on data file management and research computing techniques are being offered. The data archive must keep pace with the rapid changes in computer technology and the applications for social science research. Especially important is the integration of microcomputer workstations into the research environment, with development toward multi-user and multi-level computing capabilities. The archive staff will be working with the Cornell Microcomputer Evaluation and Development Facility in the application of microcomputers in social science research.

The data archive also hopes to develop cooperative agreements with the Cornell University Libraries in efforts to integrate services and encourage extended participation in computerized statistical information services. The data archivist participates in the development of an integrated library system at Cornell, and is a member of a cross-campus Working Group on Statistical Data Resources.

Finally, the archive staff will be active in the development of a demographic, economic, and social computerized information system on New York State, to support information, research, and training activities.
The Care and Feeding of Magnetic Tapes

Patricia A. Reslock
Center for Naval Analyses
Alexandria, Virginia

There are many obstacles to successfully acquiring machine-readable data. Difficulties usually occur during the physical transfer of the data from one site to another, the most common method being magnetic tape. Yet, even with higher quality tape and more efficient tape drives, tape processing is still plagued with problems. Parity errors, incompatible tape labels, multiple encoding schemes, and poor quality media have all contributed to tape users' frustration. In an attempt to educate the nontechnical professional, this paper will give an overview of the technical aspects of magnetic tapes and their use. Topics include the physical aspects of magnetic tape, how data is stored on tape, recording modes, tape errors, and their prevention.

A data archivist faces two main tasks: data acquisition and data maintenance. To acquire data you must know how data is stored on tape and what your installation can handle. The information necessary to successfully transfer data between sites will be covered below, as well housekeeping measures essential to keeping magnetic tapes readable.

Data Acquisition

The physical aspects of magnetic tape. A standard magnetic tape has a 10½-inch hub filled with ⅛-inch wide tape. The tape should be 2400 feet +50 feet, -0 feet. It is important to bear this in mind when purchasing new tapes, as tapes of less than 2400 feet are indicative of poor quality control.

A magnetic tape is composed of three layers. The first layer is made of oxide; this is where the magnetized data is stored. This layer must be smooth and of uniform thickness (0.00045 of an inch). The second layer (the binder) is a glue used to make the oxide adhere to the backing. The binder must be flexible enough to reduce oxide chipping, yet, not too sticky, or the tape will stick to itself. The last layer is the backing, and it is usually made of mylar. Creating a magnetic tape is a complex task; hence it is important to have rigid quality control standards.

There are three important points to be aware of when exchanging data on tape: 1) Is the tape labeled or unlabeled? 2) How many data files are on the tape? and 3) How many tape reels does the data span?

Tapes can come with or without labels, and files can come as a single file on a single tape reel, multiple files on a reel, a single file on multiple reels, or multiple files on multiple reels (reference (1),(2)). You must know what types of tapes your installation can handle and what types of utilities are available at your site for processing these types of tapes. Check to see if one form is easier to handle than another. If so, see if you can receive the data from the sending organization in that form.

Here is a list of the other items you need to know when receiving a tape:

1. Is the tape 9-track or 7-track (7-track is fairly old technology, but does still exist)?
2. How large is the blocksize of the data? Some systems cannot handle blocked data.

3. At what density was the data recorded (6250 BPI, 1600 BPI, 800 BPI)?

4. At what parity was the data recorded (even or odd)?

5. What is the character set in which the data was recorded?

Some character set possibilities include:

- ASCII American National Standard Code for Information Interchange. Most commonly found on Digital Equipment Corporation systems;

- EBCDIC Extended Binary Coded Decimal Interchange Code. IBM, Amdahl, Burroughs;

- BCD Binary Coded Decimal Character Code. CDC6600, Honeywell;


The main point to all this is: Know what your system has and what it can handle. Your Computing Services department should have a handout telling the easiest way to receive data. This information is imperative for a successful transfer of data.

I have found the government form "Transmittal Form for Describing Computer Magnetic Tape File Properties" to be an invaluable reference when receiving or sending data (reference (3)). You can use this form in one of two ways. When receiving data, fill it out the way you need to receive the tape, and send the form along with your data request. When sending data, send the form filled out with the specifications of the way the tape was created.

Tape Maintenance

How to minimize, correct, and prevent errors. Most tape errors are due to contamination of the tape, physical mishandling, or problems with tape drives and/or tape cleaning equipment. Here are some items that you want to consider to keep your tapes readable (reference (4)).

Reduce contamination. Most tape contamination comes from the tapes themselves. Oxide chips off the surface of the tape each time the tape is used. One way to reduce this chipping is to buy high quality tapes made with good binder. Other contamination comes from carelessness in handling tapes or a dirty computer room. Some ways to reduce contamination are:

1. Regularly clean tape drives. (At CNA we clean the drives at the start of every shift and before tape intensive jobs.)

2. Maintain the proper recommended temperature and humidity--this helps to reduce the oxide chipping.

3. When cleaning the floors around tapes, clean the entire floor with a damp mop--DO NOT sweep, dry mop, or dust.

4. Minimize floor waxing--if you must wax, machine buff to remove the excess wax, damp mop with cold water to harden the surface, and buff again when dry. NEVER use steel wool or other metal abrasives for buffing.

Another way to reduce contamination is to regularly clean tapes (once every eight uses). Be very careful if you decide to do this, as some tape cleaners do more harm than good. I would not recommend using a tape cleaner on a tape that is error free.

Tape drives can also produce tape errors. Your organization should have a regular tape-drive maintenance program where a field engineer does routine cleaning and alignment of the drives. This preventive maintenance is invaluable.
Long-term storage. Once a tape has been hanging on a rack for about six months, it starts to deteriorate. Pieces of dirt and chipped oxide start to cause dents in the backing of the tape. The tape may become unreadable because the dent is causing the tape to be too far away from the read head of the tape drive. The best way to prevent this type of problem is to spin tapes every six months. The preferred method is to have a program that scans the data on the tape to be sure the tape can be read. If there are problems, you will know it early and will be able to copy the data to another reel. If you wait for a long period of time to spin a tape, the damage could be permanent. The concept is the same as walking with a rock in your shoe. If you walk a block and take it out, you probably will not suffer any long-term damage. If you walk a mile with a rock in your shoe, you will have a hole in your foot.

Tape handling. Tips for tape handling include:

1. Keep tapes in the computer room.
2. Tapes should not be laid on top of the tape unit.
3. External tape labels should be sticky labels that peel off and leave no residue.
4. Never allow the beginning of the tape to trail on the floor.
5. Smoking should not be permitted around tapes.

Tape storage. Tapes should be stored in an upright position in a cabinet or shelf elevated from the floor, and as far away from sources of paper and card dust (line printers and card reader/punch) as possible. Backups of valuable data sets should be kept off site. To save money, look for a sister institution to swap tapes with instead of paying for vault storage.

Tape Transmittal. Very often tapes are damaged in the mail or while being hand carried. Tapes sent through the mail should be clearly marked MAGNETIC TAPE--KEEP AWAY FROM ELECTRIC MOTORS, SCANNING DEVICES AND MAGNETS--DO NOT X-RAY. Be sure to give the above advice to the courier for tapes being hand carried.

Conclusions and Recommendations

To ease data transfer, know the answers to all of the questions on a form like "Transmittal Form for Describing Computer Magnetic Tape File Properties" (reference (3)). Be sure to ask for some type of a dump or map of the tape being created. This makes the sending installation look at the tape after it is written—that is, it forces them to be sure something was written to the tape. Finally, be sure to get documentation on the data. These items include record layout, descriptions of all the variables and how they were derived, and the count of the total number of records in each file.

To insure in-house data reliability, scan data tapes every six months to a year to be sure they are readable. At first sign of trouble, recopy the data. Recopy valuable data sets to NEW tapes every three or four years. Try to maintain cleanliness at your site. DO NOT use a cheap tape cleaner. It will do more damage than good. Most important of all: buy the best quality tapes you can afford.

References


Profile

Rutgers, the State University of New Jersey, which was chartered as Queen's College in 1766, and was designated a state university in 1945, is a large university offering a variety of learning environments. Today the University has over 47,000 students enrolled in six separate colleges on the four campuses in New Brunswick, and the campuses at Newark and Camden each of which is over 50 miles away from the New Brunswick site. There are 24 instructional divisions and about 16 affiliated research units.

Facilities and support for academic computing are managed by the Center for Computer and Information Services (CCIS) which provides services to students and faculty who use computing for instruction and research purposes. Services include non-credit education courses, a reference center, a newsletter, maintenance of terminals and remote job entry facilities, equipment loaned to classrooms, program packages support, data archives and data base management, system programming, documentation, accounting and billing.

In 1971, the Princeton-Rutgers Census Data Project came into being through the combined efforts and finances of both universities. Since there had already existed a tradition of cooperation between the two universities on special data collections, they decided to share the purchase of 1980 Census data jointly. The project was organized with the support of the Center for Research Libraries and financial contributions from the libraries of Princeton and Rutgers as well as interested departments on both campuses. Through START-1 program, the Data Use and Access Laboratory (DUALabs), a non-profit organization, purchased the 1970 Census tapes as they became available from the Census Bureau, processed the tapes, condensed the data, and sold copies at a reduced cost to its members. Along with data modifications, several computer programs, known as the MOD series, were developed to access the data and were installed at Rutgers and Princeton Universities. All the tapes were stored at Princeton, and Rutgers copied only those pertinent to its researchers. In order to administer the project, both universities were responsible for publicity, training, and physical tape maintenance.

The project became self-sustaining by charging outside organizations for programming fees and computer cost which then covered the purchases of new tapes. The project has continued to promote collaborative efforts of cooperation and support between the two universities.

ICPSR and ROPER Memberships

In 1966, a member of the Political Science department requested membership in ICPSR. A few years later, membership in ROPER was established by Political Science department and then transferred to Sociology. As the membership in these organizations became known, an increasing number of researchers were discovering that the data produced by national archives have intrinsic research and academic value. As interest in these memberships increased, it became apparent that the individual

14
departments could not handle the workload. Since the Princeton-Rutgers Census Data Project had been functioning successfully, the CCIS decided to centralize other data bases in the same manner. The library assumed the operational control of transferring all the relevant information and materials from the individual departments and of developing administrative and ordering procedures to facilitate the acquisition of data.

Data Base Advisory Committee (DBAC)

As to the administration of the ROPER and ICPSR memberships, the CCIS favored the creation of the Data Base Advisory Committee to insure adequate communication between the CCIS, the libraries, and the departments, to determine University policy concerning future data acquisitions and to select official representation to the ICPSR and ROPER organizations. The Committee, established by the Director of the CCIS, consisted of representatives from the CCIS, the Library and the political and social science departments on the New Brunswick campus. After the first meeting, it was expanded to include representatives from the Camden and Newark campuses, also. When a discussion of the budget for the ROPER membership led to a joint membership by the Rutgers and Princeton Libraries, a representative from Princeton joined the committee. Although the committee is limited to a maximum of ten members, guests are invited and welcome. This committee, which convenes one or twice a year, discusses allocation of available resources in the departments, decides who shall represent Rutgers at the ICPSR Conference, and awards any scholarship to ICPSR science programs that become available. Communications by mail and phone are conducted continually with committee members on relevant data matters as the need arises.

The New Jersey State Data Center

In anticipation of the large amounts of data produced by the 1980 decennial census, the U.S. Bureau of the Census established a State Data Center Program throughout the country to improve access to and use of census data products. Rutgers University, as a primary participant of the New Jersey State Data Center (NJSDC), documents, distributes, and publicizes these materials. The CCIS has also made available the Census Software Package (CENSPAC), which is an all purpose statistical and retrieval program created by U.S. Bureau of Census to be utilized with the Census data.

Funding

The data activities fall within the Applications Group of the Center for Computer and Information Services which provides the facilities and the support services for academic (instructional and research) computer users. No salary lines are designated specifically for the data archives. Our programmers are responsible for computer expertise on our software and hardware for all our computer systems.

Travel requests are considered on an individual basis depending upon the overall requests for travel within the budget limitations. This fiscal year, I felt very fortunate to attend the State Data Center conference, the Association of Public Data Users, and this meeting of the International Association for Social Science Information Service and Technology. But our staff participation in such events varies from year to year.

The Rutgers membership in the Inter-University Consortium for Political and Social Research and the Rutgers-Princeton joint membership in the Roper Center are financed through the library budget. If some departments request the purchase of data outside of these memberships, the computer center coordinates the search for funding it.

Overhead expenses for office space, secretarial staff, mail, postage, telephone, etc., are not being considered here because these were already in existence when data archiving activity
came into being. The cost and maintenance of the computer equipment and cost of data processing come out of our current operating budget.

Although most of the computing with the machine-readable data is used on our IBM mainframe, some is also utilized on the VAX 11/780, which has SPSS and SCSS, and the DEC 2060 which stores a CITIBASE data file. The various departments are allotted a specific dollar amount for computing time which is based on previous year's usage and future estimates of need.

Staffing

At the time of the implementation of the Rutgers-Princeton Census Data Project, a half-time programmer analyst line was created to carry it out. When all the data activities were centralized at the computer center, the responsibilities were expanded and half-time of another programmer position was included. Unfortunately, this past year, because of many changes in personnel and the addition of a new computer, we lost ground in this area. Now less than one full-time line, shared among three people, is devoted to machine-readable data file activities. And it is not enough. The first nine months of this academic year about 250 consultations were recorded or two or three requests on an average daily basis. This figure does not include quick references in the libraries or computer-related problems which may go to the statistician or Aid Station. (There are Aid Stations on each of the campuses which are staffed by students and the computer center staff to aid in debugging all user problems.)

Sources of Data

During this academic year we have serviced more than 23 different departments on campus. Their data requests have referred to many different studies in many different fields. Requests for our census service are just as likely to come from outside the University, particularly non-profit county and state agencies as from within the University. The level of sophistication in handling MRDF's ranges from zilch to familiarity with statistical packages on the computer. All manner of problems come to the CCIS Aid Stations, the statistical consultants, and our staff during any given day. In general, the procedure in handling inquiries is fairly routine. First, we check our Rutgers University Guide to Machine-Readable Data Files to see if the file requested is already on campus. If not, the catalogs of ICPSR, the Roper Center, and the Bureau of the Census are searched for the particular file or subject requested. Data from the first two are easily obtained because of the memberships we maintain with these groups. The census inquiries require a different approach. Requestors are directed first to the printed reports. If the information is available only on tape, the researcher is assisted in ascertaining what tape contains the needed data, what census geographic area will most suit the needs of the project, and which program should be utilized. If the data needed is from a source which requires a cash outlay, the staff assists the researchers in finding funding, if at all possible.

It is difficult to determine which files are heavily used. The number of tape mounts does not give an accurate picture of how frequently the data is accessed. Most users, after accessing the tape once or twice, create a subfile on their own and continue their analytic studies on the smaller file. The sophisticated users know how to find out the tape information without checking with us. At the present time, the most frequently used studies appear to be the STF 3A tape from the 1980 Census of Population and Housing, the NORC General Social Surveys, the American National Election Studies from Michigan, and the National Longitudinal Studies from Ohio State.
Dissemination of Information

Training/Workshops/Seminars. CCIS is always looking for ways to reach more Rutgers users. In the beginning of each semester, two session seminars are conducted on familiarizing the researchers with the content of 1980 Census of Population and Housing and other Census products. Another class is held on the general Data Archives to describe the types of data available for research and study purposes. Special seminars or workshops are conducted at the request of individual units within the university and are tailored to their particular interests and needs. The close association with the reference librarians is reflected in a special seminar for the Reference Special Interest Group on the resources at the CCIS with special emphasis on the 1980 Census.

Publications/Articles/Documents. Articles on data-related information appear regularly in the CCIS bi-monthly Newsletter, but CCIS also publishes a number of Technical Documents pertaining to machine-readable data files and the computer programs available to access them. Our publications are all geared towards making data use easier for the University community. As an example of this type of publication, information was extracted from the Master Area Reference File for the 1980 Census of Population and Housing and sent to the reference librarians in all the libraries on all the Rutgers campuses. This computer output included not only the census geographic codes and corresponding area names, but also included an index by countries, an explanation of the symbolic codes and total counts for population, housing, and families. Complementing this will be our output, probably on microfiche, on county and MCD by zip code for distribution to the libraries. Finally, our most important publication is the previously mentioned Rutgers University Guide to Machine-Readable Data Files, an index of all machine-readable data files on campus.

Consultation Services. For guidance and assistance in using any of the machine-readable data, the CCIS offers a consultation service, free-of-charge, to direct users of the data, to help users select the computer program best suited for the user's need, to provide necessary program and technical documentation, and to assist in analyzing computer error messages should they occur.

Computer Reference Center. Under the auspices of the CCIS is the Computer Reference Center (CRC), a library of computer-related materials. All the codebooks, manuals, and reference books are located in the Data Archive Corner to facilitate accessibility to the widest possible range of users. These codebooks themselves can be useful tools in data analysis, sometimes eliminating the need to access the files by computer. In addition, catalogs of data holdings of several institutions which collect and disseminate data are available as well as computer-related periodicals. An information specialist who maintains and updates these materials assists users in finding the information they need.

Future Developments

Our future plans center on professional development for the staff, improvement of our excellent Census service, and expansion of the contract programming activity. In addition to in-house training workshops, staff are encouraged whenever possible to attend conferences and workshops which enhance their professional skills. It is hoped that more money can be made available for such attendance in the future. Along with attendance at these functions, staff are also urged to participate in the related organizations which sponsor these meetings, such as APDU or IASSIST. These organizations do invaluable work in fostering increased awareness of MRDF's both on-campus and in the wider academic world. Such participation

- continued on page 22
Utilities are the tools that allow the data archive to properly archive and maintain the contents of their holdings. They are the programs we use to survey or verify a tapes content, to copy tapes or print records, and to resolve tape problems. Many archives will have their own set of tools but for the new archivist there is much to learn. It is hoped this paper will encourage the new archivist to look beyond their own office for tools.

Sometimes, even the experienced archive staff are not aware of the software packages available in their computer centers. Often statistical data management packages have useful utilities the archive might use. This article will describe the utilities used most frequently at The Rand Corporation's Data Facility. A few hypothetical cases may clarify what utilities are available and will also illustrate the use of some of the utilities for the new archivist.

Also in this article, many of the examples used will be SAS programs. The reason for this is the programs do the job and they are available at Rand. I encourage the reader to make enquiries about similar kinds of software available to them.

EXAMPLE #1. Tape(s) Received for Archiving from an Outside Source

In this example our hypothetical tape has just arrived on our desk for archiving. The worse possible situation which might occur would be that the tape itself has absolutely no information about its contents nor is it accompanied by any documentation. This occurs more frequently when project personnel order a tape or data over the phone and when the tape arrives they bring it to the archive for assistance with accessing the tape. Or in another case the tape labels indicate one type of contents and accompanying documents differ in one or two of the necessary parameters. Hopefully the tape will be fully documented.

TAPEMAP (or survey or scan): This utility will read the tape looking for labels. If labels are found then the information on the labels is printed out: if no labels are found then the blocksize and density of the tape are sensed by the computer and that information is given.

At Rand and probably in most computer centers there is more than one TAPEMAP utility. Usually the utility to use is dictated by the problem and the tape involved. SAS also has a mapping program with the inappropriate name of "SAS Label." It also provides the actual footage of tape used. Following is an example of the JCL for a SAS map and a copy of the output. (This output has been edited slightly to accommodate the narrow page.)
Input:
//D0004SAS JOB (2082,30,393), 'SAS MAP',CLASS=S
// EXEC SAS,OPTIONS='N0NEWS,S=72'
//TAPE1 DD UNIT=HIGH9,
// VOL=SER=004670,
// DIS=OLD
PROC TAPELABEL DDNAME=TAPE1;

Output:
CONTENTS OF TAPE VOLUME - 004670
FILE NUMBER DSNAME RECFM LRECL BLKSIZE BLOCK COUNT FEET CREATED EXPIRES JOB
1 STF3A.0R41 FB 2016 24192 2730 1018.2 01MAY84 0000000 D0004

EXAMPLE #2.
The Tape Received Has a Data Check or Cannot Be Read by Any Mapping Utility

To resolve this problem we use a utility called Fast Analysis of Tape and Recovery. This utility reads everything on the tape. When it cannot read a section the tape is run back and forth over the drive, cleaning any specks of dust, etc., and then provides an output on the condition of the tape.

EXAMPLE #3. Tape Cannot Be Accessed by the Usual Methods

Very often the resolution to this problem is simple. The utilities we use frequently expect to find data (most utilities read a label as data) at the beginning of a tape. Infrequently we receive tapes with a tape mark at the beginning of a tape and the data is the second 'file' on a tape. Most utilities would not get past the tape mark. The resolution is to try the utility again and 'read' the second file on a tape. Instead of "label=(.blp)" use "label=(2,blp)" for jobs requiring JCL.

EXAMPLE 4. Contents of the Data to Be Verified and Tape Copied

Once we have verified the tape has the properties we expected or we have learned the true specifications the next step is to verify the contents. Do the records have the right format? Are the codes found in the records correct? If the initial map information appears to be okay it may be a good idea to copy the tape now. Checking records, etc., is often easier with tapes which are formatted to in-house specifications.

To print records: Usually we use an online utility which sets up the JCL required for a software package called DYLAKOR. With this utility we provide the tape number, the label where the data resides and the dataset name. The program works just as well if the tape is nonlabeled.

As mentioned previously, we often make use of SAS in our archive management. The following program is useful for printing records from any tape (the tapes do not have to be in SAS format.) The program does require that one variable
on the tape be identified and the location of the field in the record where the variable on the tape be identified and the location of the field in the record where the variable can be found. Following is an example of a SAS program which will print out a specified number of records (OBS=100 in this case.)

//JOB CARD
//EXEC SAS
//TAPE DD DSN=data.set.name,
// VOL=SER=volser,UNIT=TAPE,DISP=OLD
DATA;
 INFILE TAPE OBS=100;
 INPUT;
 LIST;

In the previous example if your tape is non-labeled (NL) give the DCB information either in the JCL or in the INFILE statement: (DCB=RECFM=recfm,LRECL=lrecl,BLKSIZE=blksize). If the DCB information is not known SAS assumes a BLKSIZE=32767, RECFM=U. The output of a non-label tape using these defaults will give you a readable dump allowing you to see what the data looks like and perhaps identifying the format.

EXAMPLE #5. Making Tape Copies

With an IBM computer and a straightforward tape the most reliable copy program is IEBGENER and this program is documented in the IBM manuals. However, as is also often the case, there are other programs which may save time and effort. With IEBGNENER each data set on a tape must be described in the JCL. When a tape has many data sets this can take considerable time. For this reason we use a number of copy utilities, some of which were written at Rand and are only useable here.

However, SAS has a nice tape copy utility which is very useful. This program should not be confused with their PROC "COPY" for copying SAS datasets. The SAS utility "TAPE COPY" may be used to copy any data. The nicest feature about this program is the ability to copy all files merely by writing in a range (i.e., files 1-7) or to copy files in a mixed order (i.e., files 1,2, 9,4,3). Following is an example of a SAS copy program:

//D0004SAS JOB (2082,200,393),'MCGEE',CLASS=S
//EXEC SAS
//VOLIN DD UNIT=HIGH9,DISP=OLD,
// VOL=SER=004743,LABEL=(,BLP)
//VOLOUT DD UNIT=HIGH9,DISP=(NEW,KEEP),
// VOL=SER=004196,LABEL=(,BLP),DCB=DEN=4
PROC TAPECOPY;
 FILES 1 8-10;

Example of Concatenating Files on More Than One Reel

Example: If the data is contained on more than one reel, but tapes are not concatenated. This type of file requires extra programming effort by a user when accessing the tapes. Concatenating the tapes when copying will save programming time later.
JCL to concatenate tapes:

```jcl
// JOB CARD
// STEPL EXEC IEBGENER
// OUT DD DSN=NEW.DATA,FILE,DISP=(NEW,KEEP)
//    UNIT=HIGH9, VOL=SER=000000, LABEL=(1,SL),
//    DCB=(RECFM=FB, LRECL=2016, BLOCKSIZE=24192, DEN=4)
// IN DD DSN=OLD.DATA1, VOL=SER=111111, LABEL=(1,SL), UNIT=HIGH9, DISP=SHR
// DD DSN=OLD.DATA2, VOL=SER=222222, LABEL=(1,SL), UNIT=AFF=IN, DISP=SHR
// DD DSN=OLD.DATA3, VOL=SER=333333, LABEL=(1,SL), UNIT=AFF=IN, DISP=SHR
```

Example of Changing Data Tape Format

Example: Tape has two extra characters (in this case blanks) at the end of the last record in each block. In this example the record format was 'F.' It was necessary to change the record format to 'U' instead of 'F' or the last block would not copy because it was a short block. Without this change when copying the project would have to program around the problem of the extra characters each time the tapes are used.

```jcl
// JOB CARD
// STEPL EXEC SAS
// TAPEIN DD DSN=OLD.DATA,DISP=(OLD,KEEP),
//    UNIT=TAPE9, VOL=SER=000000, LABEL=(,BLP),
//    DCB=(RECFM=U, LRECL=2872, BLOCKSIZE=2872)
// TAPEOUT DD DSN=NEW.DATA,DISP=(NEW,KEEP),
//    UNIT=HIGH9, VOL=SER=111111, LABEL=(1,SL), DCB=DEN=4
DATA NULL;
INFILE TAPEIN START=S LENGTH=L
INPUT;
S=1; L=2870;
FILE TAPEOUT LRECL=410 BLKSIZE=28700;
PUT _INFILE_
```

"Finding" Programs

The new archivist should always be on the lookout for programs to use in the archive. Following is a short PL1 program "found" at Rand and used by the archive staff as needed. This nice little program was written to "clean" the end of a tape. Sometimes we think we can write one more data set to the end of a tape only to find out we have a short reel and only part of a data file was written. Rather than have a tape with partial data we use this program to write a new tape mark, wiping out the partial data set. To use this program is is necessary to know the next file number (label number) after the last trailer label for a complete data file (in this example (4, blp)).
Data Service in a Computer Center

also often leads to use by individuals, agencies, and private firms of our contract programming, an activity that has expanded to the point that extra staffing is needed. One of our goals for the future is to be able to expand staff to handle more of these of these assignments.

Although a great deal has been done with the Census at Rutgers, more Census activity is planned. We intend to develop mapping techniques using Census data and streamline the Census programming methods currently in use. Because of the tremendous interest in Census data, plans are being made to increase the support to microcomputer users who wish to analyze Census data. The complexity of these files make subsetting time-consuming, but it is an aspect of our work increasingly in demand. All of these activities need the publicity necessary to acquaint the University community with them. Although our many workshops, seminars, and publications are widely distributed at Rutgers, still greater efforts will be made in the future.

Care and Feeding of Magnetic Tapes

-continued from page 11


For general reading:

Like the traditional library, the data archive performs many different functions to meet the needs of users. These functions include data acquisition and cleaning, development of conventions and standards for description of the data, data processing and analysis, dissemination of information about the data, storage and maintenance of data tapes, development of an inventory system and inventory controls as well as a data retrieval system, diffusion of the data, training for archive users and program development.

Data Acquisition

Obtaining new materials for archive holdings from some continuing sources of supply requires establishment of both formal and informal arrangements with institutions, departments or bureaus that produce data on a regular basis in order to obtain some or all of their productions. It is also necessary to establish priorities for the kind of data to be acquired. Since the cost of processing and maintaining a data set is often greater than the cost of acquisition, selection must be made with great care. Ephemeral or frequently replicated data sets should be acquired only when there is a concrete need for them since there is a high probability of being able to obtain popular data sets elsewhere if a local need develops. The cost of acquiring, cleaning, indexing, and maintaining a data set should be considered in relation to:

(a) the likelihood of there being multiple users;
(b) the possibility of acquiring at a later date if the need should arise;
(c) its availability at a reasonable cost and with little delay from some other source;
(d) the amount of overlap with the existing collection;
(e) the intrinsic significance of the data.

The form in which data arrives varies from supplier to supplier and from study to study. This can result in a great amount of time being spent figuring out just what it is you have received. The ultimate answer is to have funding sources or institutions conducting the survey require that arrangements for archiving be made prior to the actual funding or conducting of the research. This way, the archive can be involved from the beginning and provide guidelines and standards for researchers. A formal way to handle this would be the development of an institutional policy on minimal standards for data to be turned over to the archive. A policy statement of this type would insure that the datasets turned over to the archive meet the criterion of methodological adequacy. It has been the case that an archive decided to pass up a study of great substantive interest which appears to have been done in such a poor manner, utilizing such sloppy and shoddy techniques of data gathering or documentation that, despite the interest of the subject matter, the data set is not worth acquiring. General archive operating policy should include a list of criteria which studies should meet if the data set is to be considered acceptable. At very minimum the following documentation should be available for each study:

-continued
-continued from page 23

(a) complete and accurate codebook or description of the data structure;
(b) description of the data format;
(c) illustrations of structure and format;
(d) total size of data set;
(e) complete and accurate description of the organization of the files for the medium in which the data is stored;
(f) precise definition for each data element;
(g) complete explanation of all codes used;
(h) sample of documents used in data gathering;
(i) description of sampling procedures employed, with intended and resultant sample size;
(j) summary of training provided field-workers and coders;
(k) description of data collection procedures;
(l) name and current address of study director.

Data Cleaning

In its most simplified form, data cleaning involves processes aimed at placing data into a format that is easily handled by computers. These processes include identifying and correcting possible discrepancies between the actual format of the data and the descriptions of that format. Many archives employ specialized staff members who do this type of data cleaning.

Development of Conventions & Standards

In order to facilitate the process of utilizing data initially prepared by others it is necessary to establish conventions for coding and standards for describing the data themselves in order to:

(a) permit combining information from different collections in some reasonable way;
(b) combine samples from different studies in order to increase the number of cases;
(c) make comparisons among data sets;
(d) facilitate later analysis.

Data Processing and Analysis

The data processing and analysis function of the archive provides for the manipulation of data for the user's purpose. This may entail the reformatting of data for use at the user's local facility or providing specially prepared subset of cases or variables rather than a simple copy. Some users may need a frequency distribution for the variables (if not provided in the codebook) or simple cross-tabulations. Other users may need more detailed statistical analyses.

Dissemination of Documentation

The most important documentation produced by the archive is the codebook describing the dataset. Archive staff also prepare abstracts of data sets for inclusion in a catalog and for advertising purposes. Production of some type of archive catalog is almost mandatory since it provides not only an in-house listing of current holdings but is also the best way for a user to browse the contents of the archive. Advertising the availability of the data can take many forms. Some archives prepare and distribute their own newsletter announcing new acquisitions while others include a special data announcement section in an existing institution newsletter. Archives should also strive to maintain a collection of published material related to the data sets in order to provide examples of how the data have been analyzed already and clarify ambiguities in the interpretation of the data.

Storage and Maintenance

Internal procedures must be established to identify the current storage location of all materials. Magnetic tapes must be stored in a controlled temperature environment and protected from magnetic flux and
physical shock. They must be recopied on a periodic basis in order to assure their continued utility and, where usage is heavy, to protect against deterioration due to machine-induced wear. (See Patricia Reslock's article for a detailed discussion of tapes.)

**Inventory**

As with any collection, it is necessary that the archive maintain a catalog or index of holdings. The archivist might consider maintaining a "public" catalog and an annotated "private" catalog with additional information. The "private" catalog would include abstracts of studies added to the collection since the last published catalog update. It is also necessary to develop an internal inventory system to keep track of the current status of all studies in the archive including studies "on order" or being processed. Other internal inventory materials would include a catalog of tapes by tape or storage number and a catalog of studies by study number.

**Retrieval**

Requests from users for access to data relating to their particular topic of interest often requires the archive to search not only its own holdings but those of other archives as well and where necessary, to obtain from other archives those materials required to serve the needs of the user. For this reason it is advisable for the archive to maintain a collection of catalogs from other archives and to become familiar with the general class of holdings at other archives. Most archivists find it helpful to maintain personal contact with other archivists through the network established by professional associations (like IASSIST) in order to facilitate the exchange of information about data holdings and to keep abreast of technological developments in this field.

**Diffusion**

The data archive specializes in copying its own collection and making it available to the user at his convenience, in the form most suitable to his purposes. However, the archive must still maintain control over access to their materials in accordance with any wishes of the original donor. For this purpose the archivist usually develops a form letter which the user signs agreeing to archive terms. Once the data have been copied, the archivist fills out a standard form to send along with the data tape which describes the files on the tape...number of files, logical record length, blocksize, number of records...as well as general tape characteristics...tracks, density, format, character set, and internal labeling. Shipping data on magnetic tape requires that the tape be adequately packaged to prevent damage in transit and labeled on the outside of the package as being a MAGNETIC TAPE with a warning to keep the package away from magnets and electric motors which could destroy the data set stored on the tape. Tapes sent through the mails are often insured for the cost of recopying the data and have a return receipt included with mailing.

**Training**

Archives can perform a training function by teaching users how to make a query; where to make a query so that the appropriate data can be obtained; how to utilize the data once it is obtained; the devices available for processing; the strategies to be employed for analysis; and the kinds of interpretations that can be made from such analysis. The archivist may wish to prepare a user manual for distribution to potential users documenting how their archive is organized and including information on archive services and locally available...

-continued on page 29
Historical Background

Baruch College, originally the business School of the City College of New York (CCNY), is, since 1968, one of the eight senior colleges in the City University of New York (CUNY). Although particularly strong in the field of business, most academic fields are represented in its curriculum with the departments organized into three schools: Business, Liberal Arts and Education. The college awards business and liberal arts undergraduate degrees as well as the MBA, several other master's degrees and a Ph.D. in business. Increased interest in data resources on campus has paralleled a new emphasis on computerization. There had always been some use of secondary data on campus. Several important machine-readable data files were already available, scattered through several different departments, unorganized and with little documentation or information available outside the particular departments which possessed the files. Use was limited to those few who knew the files existed and who knew how to access them. Under the leadership of Professor Thomas V. Atkins, Deputy Chairman for Library Instruction Services, the library was able to effectively convince the college administration that data files should be conceived of as basically an information resource and, as such, the college library was the natural place for a data service.

In the Spring of 1981, a Data Archives Service (DAS) was established as part of the library's information services. The library administration added not only a data library but created an educational program whose function was to inform and instruct the Baruch community about the use of secondary data as an information resource. Because of its instructional orientation DAS was made part of the Library Instruction Services Division and was intended to complement services already provided by groups on campus such as the Educational Computer Center, the Statistics Lab, etc. Membership in ICPSR and the Roper Center were begun immediately. Data from sources other than ICPSR or Roper was purchased on an extremely selective basis due to budget restrictions. A reference collection of manuals and data catalogs was set up for use with the growing tape collection, for the first time centralizing to some extent the documentation for both mainframe software and secondary data. The training programs were begun, on a limited basis, almost immediately.

Almost at once, Baruch College began to lobby for a University-wide ICPSR membership. There was some precedent for this since there had been a previous City University ICPSR membership which had lapsed due to administrative problems. Baruch's efforts joined those of a significant number of faculty and administrators at other CUNY colleges who had long been interested in seeing a return of the university-wide membership. Finally, in July of 1983, these combined efforts were successful and the senior colleges of the City University of New York became a federated member of ICPSR, one of the largest federations
in the Consortium. To begin with the federation included only the four-year institutions. It is assumed that the two-year colleges will join at a later time if there is sufficient interest. To encourage the success of the federation, Baruch's administration, both college and library, willingly accepted the college's appointment as coordinator for the new membership.

Funding

As is often the case with academic institutions, there was little additional funding available. When the service began it operated on the proverbial "shoestring" with support from the library budget and using library personnel. Some additional financial assistance came from the Title III grant awarded to Dr. Atkins for the development of a Graduate Business Resource and Study Center. Sufficient money for computer use and tape storage were allocated from the general research funds of the Baruch College Educational Computer Center.

When Baruch became the coordinator of the City University ICPSR membership, the University Chancellor's Office paid the Federation's membership fee. Half-time services of Baruch's data archivist, additional student assistance, and some money for non-personnel expenses such as documentation, magnetic tapes, supplies, software and travel funds were funded by additional support from the Chancellor's Office and members of the ICPSR Federation. Since Baruch contributed its facilities and the services of its already established data library, it was not required to contribute further funds. University support is limited to expenditures associated with the ICPSR Federation while Baruch uses its own funds for purchase of non-ICPSR data, special equipment and its own data services.

Staffing

Baruch's data archivist is assigned part-time to the Baruch service and part-time to the CUNY Center which is also staffed by a part-time graduate assistant and undergraduate student assistants. The archivist, a trained librarian, set up the data library, organized the tape collection, developed the documentation collection of appropriate codebooks and manuals, and established the research consultation service. Once organized, basic maintenance of the tape and reference collection have been assumed by the graduate assistant, who also provides assistance with the development of educational programs. As the service expands, it is expected that additional graduate assistance will be needed. Undergraduate students assist with clerical duties, as does the secretarial staff of the Library Instruction Division. The staffing is based on an assumption that computer and statistical consultation is available from Baruch College's Educational Computer Center or, in the case of other CUNY users, at their home campuses.

Equipment

The City University has a large central computer installation (CUNY/UCC) used by all the colleges as their main facility. The hardware at the UCC includes an IBM 3081, an IBM 3033, and an Amdahl 470/V6-II. In addition to the mainframes, there are high speed printers, graphics equipment and software installations of most of the major statistical packages. Almost all the senior colleges, including Baruch, have supplementary equipment including mainframes, minis and microcomputer labs. CDA stores copies of its tapes at the CUNY UCC on a permanent basis so that they are easily available to all campuses. For convenience Baruch facilities are often used for small printing jobs since the CUNY UCC is located some 50 blocks to the northwest.

The Center itself has had only a Decwriter 300 band printing terminal for maintenance of its tape collection and development of on-line demonstrations for seminars. Recently a Volcker Craig CRT and an IBM PC XT were received.
This equipment will be used for present activities of the Center in addition to development of instruction in and assistance with microcomputer data analysis, an area in which the CDA intends to specialize. For training seminars which include on-line demonstrations of secondary data resources, the Center has had access to the Baruch College Graduate Business Study and Resource Center seminar room which is equipped with several Decwriters for workshop participants and an Electrohome Projector which projects an enlarged image from a video terminal.

Physical Environment

If the Data Archives Service at Baruch, or for that matter, the CUNY Data Service had waited for proper housing, it would not exist today. Space is at a premium almost everywhere in the University, but nowhere more than at Baruch College. The data library was begun in one of the faculty offices of the Library Instruction Division which at that time housed three other faculty members in addition to the part-time data archivist. At this writing, a separate room serves as office space for the archivist, as a workroom for maintenance of the data collection, a storage room for the documentation collection, and as consultation space. There is some additional storage space for the master tape copies. Presently, plans are being made to acquire additional space which will serve as a workroom for the students in maintaining the tape collection and the data archives files. The original office will then be freed for use as office space and for data consultations. On the desiderata list is a terminal room for users which will encourage data use in the data library and allow for on-line data consultations.

Dissemination of Information

When the Baruch Data Archives Search Service became the coordinator for the CUNY ICPSR membership it expanded upon its own primary emphasis, the education of faculty and graduate students, to bringing about a general awareness of the potential of secondary data analysis and ICPSR files in particular. A monthly annotated list of ICPSR data available to all CUNY faculty is mailed to the campus liaisons, the libraries and specific data users. At Baruch, the CUNY list is supplemented with a listing of datasets available to only Baruch faculty and students. A Baruch data directory is in preparation which will list and index by subject all data available on campus including nonbibliographic databases accessed through Computer Search Services. These general listings are supplemented by data bibliographies on specific topics which are prepared for seminars and then mailed on request. The semi-annual newsletter published by the Baruch Graduate Business Study and Resource Center and mailed to all departments has carried a section on the Baruch Data Archives since its inception. But timely information and publicity for the entire CUNY community remains a difficult problem because staff is limited and the community to be served is large, disparate and separated by sizeable distances.

For Baruch and CUNY the most important methods of reaching out to both new and sophisticated data users has been the training program and the seminar series. Each seminar focuses on the information resources for the study and teaching of a particular inter-disciplinary topic. Each seminar includes some mention of bibliographic databases available on the topic as well as a brief overview of information sources in print. The greatest portion of the two hours, however, is devoted to machine-readable data files, particularly those from ICPSR. Where appropriate, government data files or other data archives which specialize in the topic are mentioned. An on-line, interactive demonstration of the contents of an important dataset in the field concludes each seminar. Topics covered have been Urban Problems, Women, Youth, and Consumer Behavior. Plans for the future include seminars in this
format on the national elections, health care, education, marketing, etc. The seminars held at Baruch have been supplemented by visits to the individual college campuses with a "What is ICPSR" format also including on-line demonstrations.

Future Developments

At this point it is expected that the CUNY ICPSR Federation will continue as an integral part of the university's information resources and Baruch will maintain its own data service as well. Plans for the future for both services are inextricably linked. First on the agenda is the solution to the problem of reaching the huge CUNY community. In some part this will be done by renewed emphasis on successful programs, but additional services will be offered as finances and personnel permit.

We intend to continue and increase training in the availability of data resources and how they are used for research and teaching. These seminars, we believe, assist in motivating faculty to enlarge their uses of data and secondary analysis both in research and in instruction. Both the interdisciplinary seminars held at Baruch and the ones held at the individual colleges will be increased. In addition, "hands-on" workshops actually using data will be held in the specially equipped Baruch on-line classrooms. Not only our training will increase, but also our services. We intend to facilitate faculty and student use of microcomputers for data analysis by purchasing or preparing subsets on diskette and supporting microcomputer statistical packages. Special workshops are planned, for example, on the use of ABC, ICPSR's instructional statistical package. As a part of this effort, special instructional packages will be developed similar to teaching packages prepared by the Library Instruction Division for bibliographic instruction. These are intended for use in our course-related lectures series.

Our informational services will expand this year with a new data directory containing tape access information, a subject index and enlarged annotations. For Baruch this will tie together all campus files, while the CUNY edition will list ICPSR data available. This directory is a preliminary step in our long-term goal of having an on-line dataset catalog. Publication of our newsletter on a regular basis and special publications on major datasets are also in our plans for future development.

DATA ARCHIVE MANAGEMENT

-statistical packages used for analysis. Another useful manual would be an internal document describing the operational procedures of the archive for use in training new staff.

Program Development

Archives can play a role in the development of new programs, particularly in the collection and creation of specialized data and by encouraging the creation of new computing power.

The author wishes to thank David Nasatir, whose article "Operational Considerations of Archives" in Howard White's Reader in Machine-Readable Social Data served as the basis for both the workshop presentation and this article.
MEMBERSHIP AND SUBSCRIPTION FEES

The International Association for Social Science Information Services and Technology (IASSIST) is a professional association of individuals who are engaged in the acquisition, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information systems specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompasses hardcopy as well as machine readable data.

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

Membership fees are:

REGULAR MEMBERSHIP: $20 per calendar year
STUDENT MEMBERSHIP: $10 per calendar year

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

INSTITUTIONAL SUBSCRIPTION: $35 per calendar year (includes one volume of the QUARTERLY)

Ms. Jacqueline McGee
The Rand Corporation
1700 Main Street
Santa Monica, California
U.S.A. 90406
(213) 393-0411
IASSIST — MEMBERSHIP APPLICATION

Please accept my application for membership in IASSIST:

NAME __________________________ PHONE ( ) _______________________

INSTITUTION AFFILIATION ___________________________________________

MAILING ADDRESS __________________________________________________

________________________________________________ ZIP/POSTAL CODE ______

I am requesting:

☐ REGULAR MEMBERSHIP $20.00
☐ STUDENT MEMBERSHIP $10.00
☐ INSTITUTIONAL MEMBERSHIP $35.00

MAKE CHECKS PAYABLE TO IASSIST

My primary interests are:

☐ ARCHIVE SERVICES/ADMINISTRATION
☐ DATA PROCESSING/DATA MANAGEMENT
☐ RESEARCH APPLICATIONS
☐ OTHER (SPECIFY) _____________________________________________

☐ PLEASE INCLUDE MY NAME IN THE IASSIST MEMBERSHIP DIRECTORY

MS. JACKIE MCGEE
THE TREASURER, IASSIST
RAND CORPORATION
1700 MAIN STREET
SANTA MONICA, CA 90406
PRESIDENT: SUE GAVREL, MACHINE READABLE ARCHIVES, PUBLIC ARCHIVES OF CANADA, 395 WELLINGTON STREET, OTTAWA, ONTARIO K1A ON4, CANADA 613/993-7772.

VICE PRESIDENT: HENK SCHRIK, STEINMETZ ARCHIVES, HERENGRACHT 410-412, 1017 BX AMSTERDAM, THE NETHERLANDS (0)20/225-061.

REGIONAL SECRETARIES

ASIA: NARESH NIJHAWAN, INDIAN COUNCIL OF SOCIAL SCIENCE RESEARCH, DATA ARCHIVE, 35 FEROZESHAH ROAD, NEW DELHI, 11001, INDIA 388-837.

AUSTRALIA: TERENCE BEED, SAMPLE SURVEY CENTRE, UNIVERSITY OF SYDNEY, SYDNEY, NEW SOUTH WALES, AUSTRALIA 2006 (02)/692-3624.

CANADA: WENDY WATKINS, SOCIAL SCIENCE DATA ARCHIVE, ROOM A715, LOEB BUILDING, CARLETON UNIVERSITY, OTTAWA, ONTARIO K1S 5B6, CANADA 613/231-7426.

EAST EUROPE: BARTLOMIEJ GASIOROWSKI, KOMITET NAUK POLITYEZNYCH, PAN DANYCH SKRYTKA POCZTOWA 12, 00-955 WARSZAWA, POLAND.

WEST EUROPE: ERIKA VON BRUNKEN, KAROLINSKA INSTITUTE LIBRARY AND INFORMATION CENTER, P.O. BOX 60201, S-104 01 STOCKHOLM, SWEDEN 46-8/232270.

UNITED STATES: JUDITH S. ROWE, COMPUTER CENTER, 87 PROSPECT STREET, PRINCETON UNIVERSITY, PRINCETON, NEW JERSEY 08544 USA, 609/452-6052.

MEMBERS-AT-LARGE

JOHN DE VRIES, DEPARTMENT OF SOCIOLOGY, CARLETON UNIVERSITY, OTTAWA, ONTARIO K1S 5B6, CANADA.

SUE A. DODD, INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE, ROOM 25, MANNING HALL 026A, UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL, NORTH CAROLINA 27514 USA 919/966-3346.

CAROLYN GEDA, INTERUNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH, P.O. BOX 1248, UNIVERSITY OF MICHIGAN, ANN ARBOR, MICHIGAN 48105 USA 313/764-5199.

JACKIE MCGEE, THE RAND CORPORATION, 1700 MAIN STREET, SANTA MONICA, CALIFORNIA 90406 USA 213/393-0411, X7351.