# A New Approach to Data Archive Decision-Making

by

Jane Wolin
Head of Programming
Center for Computer and Information Services
Rutgers, the State University
Hill Center, Busch Campus, P.O.Box 879
Piscataway, N.J. 08854
(201) 932-2483

## ABSTRACT

As is the case at many educational institutions, Rutgers University is a member of several organizations from whom machine-readable data files are acquired. Up until last year, the administration of these memberships was handled by faculty members of various departments. However, since this proved to be an unsatistfactory arrangement, representatives from the interested departments, from the CCIS (Center for Computer and Information Services) and from the Rutgers Library met and decided that all data archives should be managed by the computer center. However, to ensure that the various departments would continue to have an active role in making the important decisions and establishing policy, a Data Base Advisory was created.

This paper will discuss the events leading up to the formation of the Data Base Advisory Committee, describe its composition, purpose and activities and report on its first year of successful operation.

## Data Archiving "Growing Pains"

Rutgers University, like many other educational institutions, has memberships in several national organizations from whom machine-readable data files are purchased; in addition computer tapes are also acquired from governmental agencies, non-profit corporations and various colleges and universities. Some of these organizations with whom Rutgers has been dealing over the years include the Inter-University Consortium for Political and Social Research (ICPSR), the ROPER Public Opinion Research Center, the U.S. Bureau of the Census, the National Bureau of Economic Research, Tri-State Regional Planning Commission, Data Use and Access Laboratories (DUALabs), the National Institute for Mental Health (NIMH), and many local and state agencies.

Data archiving began at Rutgers University early in this decade at the academic computer center, CCIS, (Center for Computer and Information Services). At that time the libraries of Rutgers and Princeton joined together and contributed funds to create the Princeton-Rutgers Census Data Project. This close cooperation between the two universities' libraries and computer centers has been extremely successful because together we have been able to acquire more data and provide better services for our users than either one would have been able to do alone.

During the time that the Princeton-Rutgers Census Data Project was establishing itself, a faculty member from the Political Science department persuaded the Rutgers Library to pay for annual memberships in ICPSR and ROPER. Since this professor was the primary user of the data files, he was designated as the official representative and served in that capacity for several years. Over a period of time however, as he became more involved in teaching and doing his own research, he found the administrative tasks more and more burdensome. To ease this situation, the ROPER membership was transferred to the Sociology Department because by this time their faculty members were the main users.

Although this writer, who managed the Princeton-Rutgers Census Data Project at the CCIS, kept in close contact with both the Political Science and Sociology departments, it became increasingly evident that having three separate groups acquiring machine-readable data files was both inefficient and impractical for several reasons:

1. There was no centralized location where faculty and researchers could call to obtain information concerning the availability of data files, titles, data set names, volume serial numbers, code books, etc.

2. Because of the commitments of the Political Science and Sociology departments to their own students and researchers, there was no time to service other users.

3. No publicity was given concerning the existence of these data, so other departments did not even know that files were on hand or could be ordered.

4. If users were lucky enough to be able to reach the official representative, there was a long delay before data were ordered.

5. No consultation services were available to assist people in using the data or to help them if problems were encountered.

Once the Princeton-Rutgers Census Data Project was well under way and the information from the 1970 Census of Population and Housing was being used extensively, there evolved a natural tendency for the two computer centers to combine efforts and provide joint funds to obtain other data such as the National Longitudinal Surveys, and the Current Population Surveys. Since Princeton had its own membership in ICPSR, data from this source was (and still is) obtained separately. On the other hand, Princeton was not a member of ROPER, but many of their faculty expressed a desire to obtain some of ROPER's files. Again, the Princeton and Rutgers Libraries with the concurrance of the computer centers reached an agreement to share these data.

As the quantity of data grew, the number of users and the demand for service also grew. Unofficially, the CCIS began to explore the possibility of finally centralizing the data archiving at the computer center because 1) the physical data were there already, 2) the CCIS Program Library was the natural place for code books and documentation to be stored, 3) the CCIS was already successfully providing this service for census data and other miscellaneous files and 4) the users were constantly frustrated when trying to obtain information and/or data from the campus ICPSR and ROPER official representatives.

Much to our surprise, everyone was in favor of the transfer. The people from political science and sociology wanted to be relieved of handling the administrative details such as ordering data, creating backup tapes, keeping track of volume serial numbers and data set names and dealing with the steadily growing number of questions and inquir-

ies. The Rutgers Library was particularly anxious for data archiving to be handled by the CCIS. Since the Library had funded the ICPSR, ROPER and census projects, they were able to see the difference in the way the three were managed. Most of all, the Rutgers and Princeton computer centers were anxious for the transfer to take place so that duplication of effort could be eliminated and all data archiving would be centralized at a single location on each campus.

There appeared to be only two problems. The official representative to ICPSR is empowered to decide who shall receive the travel funds to attend the ICPSR Summer program and who shall represent the university at the ICPSR biannual meeting. Since the Political Science department always had had control, they did not want to give up this role entirely. The second problem dealt with the fact that under our contract with ROPER, only a limited number of files could be ordered. If contention for these units arose, who would decide which requests would be honored?

In an attempt to circumvent these two stumbling blocks, it was suggested that an advisory committee be formed which would have representatives from the CCIS, the Library, each of the interested departments and a representative from the Princeton Computing Center who would vote only on matters concerning data holdings held jointly (e.g., ROPER, census, etc.). To this end, a planning meeting was held in March of 1976. The CCIS presented a proposal defining the composition and the functions of the committee and outlining the services which the CCIS would perform. These proposals were discussed, amended and revised to the satisfaction of all, and the Data Base Advisory Committee was conceived!

## The Birth of the Data Base Advisory Committee

The basic components of the agreement reached during the Spring of 1976 are as follows:

1.  The CCIS will provide the following services:

    a.  Serve as official representative to ICPSR and ROPER.
    b.  Process all requests for data.
    c.  Handle the ordering of all data files.

d. Maintain tape security.
e. Publish a list of all data holdings.
f. Prepare annual reports to the Library and the Data Base Advisory Committee.
g. Forward all communications from ICPSR and ROPER of an academic interest to the appropriate department representative.
h. Report to the Advisory Committee once a year or more often if necessary.

2. The CCIS director shall establish a Data Base Advisory Committee as follows:

a. Members of the Committee shall be appointed by the Director of CCIS and shall be composed of a representative from CCIS, the University Library and those academic disciplines which use the data.
b. This Committee will have a minimum of 5 members and a maximum of 10.
c. The Committee will determine the allocation of available resources to the departments.
d. The Committee will decide who shall represent Rutgers at the biannual ICPSR conference.
e. The Committee will award any scholarships to the ICPSR summer program.
f. Princeton shall appoint a member to the Committee who will vote on matters concerning joint data holdings.

During the summer of 1976 and continuing into the Fall, the actual transfer of responsibilities from Political Science and Sociology to the CCIS progressed. Department chairpersons on each of the Rutgers campuses, Camden, Newark, and New Brunswick, were contacted requesting the names of faculty members who were particularly interested in using machine-readable data files. The Director of the CCIS then wrote to these people requesting that they serve on the Data Base Advisory Committee. By November, all administration of data archiving was being done by the CCIS, and the members of the Committee had finally been selected. The first meeting was scheduled for the first week in December, and a proposed agenda was mailed to committee members in advance to insure that all topics of mutual interest would be discussed.

At this meeting, the Director of CCIS officially welcomed everyone, and the members were introduced to one another. The Committee consists of two representatives from the Library, a political scientist from Rutgers, New Brunswick, one sociologist from Rutgers, Camden, and one from Rutgers, Newark, a liaison from the CCIS to the New Brunswick Sociology department, a member of the Princeton Computing Center and this writer representing the CCIS.

After announcements were made concerning recent data acquisitions, CCIS presented its preliminary version of a computerized listing of all data holdings. It had taken almost six months to compile, and it contained a cross-reference listing by title, subject and author. This was the first attempt at providing users with a centralized reference guide. The committee was unanimous in its praise and immediately decided to entitle it, "The Rutgers Guide to Machine-Readable Data Files." During the ensuing discussion and review of the "Guide", the Committee made many useful suggestions for improving its contents, recommended that copies be sent to the reference librarians at each of the libraries around the university and determined that updates to the "Guide" be published in the CCIS Newsletter. A new version of the "Guide" will be disseminated every year. Princeton announced that since they too were in the process of creating a new, computerized listing of their data holdings, they would conform to the Committee recommendation and entitle their publication, "The Princeton Guide to Machine-Readable Data Files."

Concerning the functioning of the Committee itself, this writer was designated as director of the Committee whose responsibilities include scheduling meetings, writing the proposed agenda and mailing it out with reminder notices, sending ICPSR and ROPER announcements of an academic nature to the interested departments, preparing and mailing out the minutes of each meeting and writing annual reports summarizing the year's activities. The Committee also decided that recommendations for adding new members will be given to the Director of CCIS should the occasion arise. However, to insure that there is the widest possible coverage of the Committee's activities, meetings will be publicized in the CCIS Newsletter, guests from the university community will be welcome and encouraged to attend, and minutes of the meetings will be sent to all people on the CCIS mailing list who indicated an interest in data bases or census.

At the conclusion of the first meeting, the members all expressed their pleasure in the way the Committee was beginning to function. The Camden and Newark representatives were particularly enthusiastic because in the past their geographic distance had prevented them from obtaining up-to-date information. They had often felt (with justification) that they were not a part of the mainstream of computer usage for instructional and research purposes. As a result of this meeting, the Camden representative requested that a special presentation be made to the faculty on that campus concerning the availability of data and services. He took care of local arrangements and publicity, and two members of the New Brunswick CCIS staff traveled to Camden in February to give an overview of the kinds of data already on site and available through ICPSR, ROPER, and the Princeton-Rutgers Census Data Project. Reference materials were described and the consultation services provided by CCIS both at their remote Camden location and at the main installation in New Brunswick were discussed.

As a result of the presentation at Camden and other similar ones held at different departments and the publicity of the Data Base Advisory Committee and "The Rutgers Guide to Machine-Readable Data Files", we were able to report at the second Committee meeting in the spring of 1977 a marked increase in usage of our data files. Naturally the Library is particularly pleased because now they feel that the university is receiving better value for the money that they have invested.

Also during the second meeting, the Committee made policy decisions concerning our ICPSR and ROPER memberships. For example, since only a limited number of data files can be purchased from ROPER as part of our annual fee, these resource units were allocated to various users in as equitable way as possible. The possibility of having the Princeton and Rutgers Libraries catalogue our machine-readable data files and/or code books was also discussed as were other procedural and political issues of mutual concern.

## A Look To The Future

At this writing, the newly structured archiving of computerized data at Rutgers is now a year old. Obviously a good part of this time was spent in organizing information, developing administrative procedures and

publicizing the availability of data files. However, from past experience, we already knew that simply making the data available was not enough. Most researchers are not sophisticated users of the computer and many are totally ignorant of how the computer can and cannot be used to assist them in their work. As a result, the formation of the Data Base Advisory Committee and the creation of the "Guide" are only preliminary steps.

As mentioned above, part of the problem of having our ICPSR and ROPER memberships administered by separate departments was the lack of user services, namely consultation, education, documentation and assistance if something goes wrong. By centralizing all data archiving at the CCIS, members of the university community now have these services available to them through the CCIS staff. Consultants will guide, advise and assist users not only in the use of the various data files, but also in related areas such as statistics and the use of statistical program packages, graphics, data management, information retrieval programs, data base management systems and many more.

In addition, every year the CCIS staff conducts non-credit courses that are open to faculty members, graduate students and if classroom space permits, undergraduates and external people who want to learn to use the computer in their work. Introductory topics as well as advanced computer concepts are taught in order to satisfy the needs of a wide range of users. The CCIS also maintains AID Stations at the central computer site and at many remote locations where people can go for help if something goes wrong with a computer run.

As important as these user services are however, there still is one more area which is extremely necessary. Part of the reorganization which took place during the last year included relocating all of the printed reference material required to use the computer files. These manuals and code books, rather than being scattered about in different people's offices, have all been placed for easy reference in the CCIS Program Library located at the computer center. However, cataloging these code books, indexes, reference guides and other publications still remains to be done. Then, after the American Library Association determines the rules for cataloging machine-readable data files, the next large task will be to catalog all of the data itself. In this way, when researchers go to the card catalog

at the main Library, not only will they
find out what books are available on
their topic, but they will also discover
that machine-readable data files exist.
It is hoped that the Princeton and
Rutgers Libraries with the help of the
computer centers will be able to work
together on this project and thus effect
a considerable cost savings.

As the reader can see, a great deal
of progress has been made at Rutgers
University in the field of data archiv-
ing, but it is an on-going process, and
much work still remains to be done.
Circumstances, situations and politics
differ from one institution to another,
so perhaps our approach may not
necessarily be applicable to other
installations.  However, because cen-
tralizing data archiving has initially
been so successful and has fostered such
close cooperation between the computer
center, various departments, the Library
and another nearby university, other
educational institutions might like to
look into the possibility of developing
similar relationships.