

# ORGANIZING A DATA LIBRARY

*Ann E. Gerken, Data Archivist  
Cornell Institute for Social and Economic Research  
Cornell University*

## 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 consortiums, 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-PC1, 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-university 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 Micro-computer 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.