

Part

4

Collection Development Policies and Procedures

Collection development policies and procedures, which are routinely used in library environments, are applicable in any setting where data files are stored for use by others. This section contains sample documents of collection development policies and statements about the conditions for depositing data in a collection. Data deposit and release agreements are important if you intend to archive data or redistribute data. These examples should be useful when writing your own policy statements defining the scope of what to include and exclude in your collection. You may also want to see the *ICPSR Data Deposit Form* which is available from their web site (<http://www.icpsr.umich.edu/ICPSR/Archive/Deposit/depositform.html>).

Collection Development Policies

Subject: Data Files for Social Research (Binghamton University)

Scope

Data files for social research are numeric computer files which are stored in a format which requires appropriate computer hardware and software to use. Data files are acquired in several ways: our membership in the Inter-university Consortium for Political and Social Research and Sociometric's SocioNet Data Consortium, our affiliation with the U.S. Bureau of the Census State Data Center Program and the Federal Depository Library Program, as well as directly by firm order or annual subscriptions from data producers. Data files may be purchased, leased, given as a gift, or acquired as a result of a program to which the University or Libraries belong.

Programs and Clientele

Our collection of data files primarily supports quantitative social research by graduate students and faculty in the Departments of Economics, Geography, History, Political Science, and Sociology; and the Schools of Management and Nursing. Other departments and units which make use of the collection include Africana Studies, Anthropology, and the School of Education and Human Development. Advanced undergraduates in several social science disciplines and the School of Management increasingly make use of these resources.

Collection Guidelines

Data files are acquired based on patron request, except in the infrequent case where it is known that an update to a particular file is desired. No time period or subject covered is outside the scope of the collection. Factors which do impact the acquisition of data are: cost, physical format, software compatibility, access restrictions, extent of target user group, and availability of documentation. Data files are acquired on various media including: magnetic tape cartridge, diskette, CD-ROM, or electronic transfer via the Internet.. Special effort is made to collect and maintain technical documentation and codebooks for the data we acquire, as well as for data not yet acquired which may be of potential interest to our clientele. Statistical software manuals and reference materials to aid in the identification and use of datasets are also collected as supporting materials. We selectively license/purchase statistical software and data management software for staff use to provide access and research services.

Other Considerations, affecting the development of this collection

A large portion of our collection of data files is stored in Computing Services. This requires cooperation with this division for the acquisition and management of data received on magnetic tape or cartridge, and via electronic transfer. Our collection of data files constitutes a data library rather than an archive, because we do not attempt to preserve items we acquire indefinitely.

Draft Collection Development Statement... (UCSD)

DRAFT COLLECTION DEVELOPMENT STATEMENT

FOR MACHINE READABLE DATA FILES

INTRODUCTION

This policy is intended to provide guidelines for the acquisition of machine readable data files. As the technology for the electronic storage of information changes, and as different kinds of information are made available in machine-readable formats, this policy will require revision. The current guidelines provide adequate physical and bibliographic access for a rather limited collection of specialized files. As the collection grows in scope and size, new guidelines may be required.

DEFINITION OF MACHINE READABLE DATA FILES

A machine readable data file (MRDF) is any information - numeric, textual, bibliographic, or some combination of these - stored in an electronic medium which is readable only by machine. A typical MRDF would consist of numeric data stored as electronically recorded signals on magnetic recording tape and readable by a computer, but other formats such as floppy disks, hard disks, video disks, and compact disks may become more common.

GOALS

The goals of this machine-readable data file policy are to a) affirm that the Library will acquire information in the format or formats which are most useful to the UCSD community, b) provide criteria for selection to MRDFs including those criteria which are specific only to MRDFs, and c) outline the role of the Library in providing access to MRDFs.

SCOPE

This policy covers data in any machine-readable format. It covers any type of library acquisition including purchase, lease, gift, or other means of acquisition which add material to the library collection.

POLICY

Machine readable data files acquired by the Library must meet the same criteria that other formats must meet to be added to the collection, i.e., the data or information must support an identifiable current or future research or curriculum need on campus. The extent of this need must justify the expenses of acquiring, processing, and maintaining the data file. In addition, acquisition of MRDFs by the Library should involve the following considerations:

1. Codebooks. There must be a complete and accurate codebook or user guide which details data structure and format, defines each data element, and explains all codes used.
2. Other Documentation. Other documentation, such as a description of how the data were collected, may be required for

some data sets in order for the data user to evaluate and use the data.

3. Physical Format. The machine-readable format must be compatible with machines available to the UCSD community.

4. Software. Before acquiring a machine-readable data file, the Library must be assured that the UCSD community will have software access to the data file. While the Library does not, at this time, have the staff to create software, software access can be assured in any of several ways:

- a. The library can acquire, through purchase or other means, available software.
- b. The data file may be accessible by software already available to the UCSD community.
- c. Programming for accessing the data file may be made available (for example, by an academic department).

5. Duplication of Data. As with other formats, the content of the data should be evaluated in terms of how much it duplicates data already available in the collection. It may be appropriate, however, to make data available in more than one format (e.g., in printed format and in machine readable format). Each case should be evaluated individually.

6. Authority of Data. As with other formats, the authority and completeness of the data should be considered.

7. Online Availability. Many MRDFs are available online. Before purchasing a file which is available online through vendors or search services, the costs and uses of on-campus availability versus online availability should be considered.

8. Accessibility. Some data files may have restrictions on their use. Examples of restrictions include: data may be limited to "academic use" only; data may not be copied; data may contain proprietary or confidential information. Data files which have restrictions with which the Library cannot comply will not be acquired.

PROCEDURE

The Library will acquire only those data files which meet the policy criteria stated above and this will normally occur only after a member of the faculty requests a specific file. The Library will evaluate the acquisition using the criteria listed above. The evaluation process will include the appropriate subject bibliographers, the MRDF bibliographer, the AUL for collection development, and the faculty in appropriate departments.

The Library will acquire codebooks, user manuals, and other documentation along with any datafiles and will catalog and maintain these materials.

THE ROLE OF THE LIBRARY

The Library has several roles in assuring campus access to machine readable data files.

1. The Library will help identify the existence of data files not in the Library collection. The Library will continue to take an active role in keeping its users informed of the availability of information pertinent to their needs and will make a particular effort to identify machine readable data.

2. The Library will acquire the data files, and will insure that the files meet the standards set forth in the selection criteria listed above. When data files are available from more than one source, the Library will determine the best and most reliable source.

3. As a federal government depository library, the Library has a particular commitment to making government-produced information available. Because the federal government does not, at this time, include machine readable formats in the depository system, the Library will take an active role in attempting to assure access to such government- produced machine readable information.

4. The Library will maintain memberships in organizations such as the Inter-University Consortium for Political and Social Research and the International Survey Library Association in order to provide reliable and easy access to datafiles. The funding sources for such memberships have not been established at this time.

5. The Library will acquire, catalog, and maintain codebooks and other user guides, making them readily available to users of the Library collections.

6. The Library will provide reference service for the printed materials and referral service to the machine readable data files as part of the routine of providing information services.

7. The Library may produce catalogs, lists, or other materials to aid in identifying machine readable data files in the Library collection.

Collection Profile 1990 (UCSD)

Subject:	Machine Readable Data Files (MRDF)
Bibliographer:	Jim Jacobs
Degrees granted:	N/A
Collection Size:	140 studies consisting of over 1300 datasets; 250 volumes of codebooks.
Locations:	Instructional Computing Center and Central University Library.
No. of students:	N/A
No. of faculty:	N/A

Brief statement of purpose and scope.

Scope: Machine Readable Data Files (MRDF). Any information that is acquired or stored in a format which requires a computer to read or use the information. This collection profile covers numeric data and text files used for Social Science research and instruction. This profile does not cover computer-stored images, cartographic information, or computer programs. Inclusion of these other types of files or other disciplines will require modification of this profile.

Purpose: To support research and instruction in those UCSD Social Science programs which require information in machine readable form. Also to support research and instruction in those areas where the information required is available only in machine-readable form.

Other: N/A

I. Description of existing collections

The collection as of the Fall quarter 1990 includes approximately 132 machine readable studies consisting of over 1300 separate computer files (datasets) on over 300 reels of computer tape. The number of tapes includes both working copies and backup copies. An additional eight studies are kept only on magnetic disk as part of the Social Sciences Data Base (SSDB) project on the Social Science Computer Facility machine "weber."

In addition to the machine-readable data files, the collection includes approximately 250 volumes of printed codebooks and other documentation. This includes codebooks for which the library does not own datasets as well as codebooks for which the library does own datasets.

II. Programmatic needs

The collection of machine readable data files primarily supports graduate and faculty research in the Departments of Economics, Political Science, and Sociology, and the Graduate School of International Relations and Pacific Studies. Other Social Science departments and the Departments of History, Psychiatry, and Public Health, and the Center for U.S./Mexican Studies also make use of this collection. Undergraduates in the Social Sciences (primarily Economics, Political Science, and Sociology) also use the collection.

The purposes of the collection are to support any program which requires information in machine-readable form and to collect that information which is both available only in machine readable form and needed for research or instruction at UCSD.

The collection of codebooks provides information about the data in the collection and about data not in the collection. It thus helps data users use the data on campus and select and identify data not on campus.

III. Collection Guidelines

A. General Considerations

For the most part the collection is being built passively in response to requests for machine readable data from faculty and graduate students. There is no active program for acquiring data in anticipation of possible future needs. The exception to this general rule is the acquisition of new editions of a few studies as they are made available. Examples include: The General Social Survey, The American National Election Study, International Financial Statistics, and the Citibase Economic Time Series. These are acquired without specific new requests for each new edition. The MRDF bibliographer monitors the use of these and other such studies to insure that the library only acquires those studies which are actually being used.

The collection consists almost exclusively of numeric data, but the Library is acquiring some textual data (notably two newsletters on Latin American affairs). Although this collection profile reflects the needs of social science research, it is possible that the Library will modify this profile to include other areas such as the humanities.

Approximately seventy-five percent of the studies acquired so far by the Library have been acquired from the Inter-University Consortium for Political and Social Research (ICPSR). The Library pays an annual membership fee to the Consortium in order to maintain the campus relationship with the Consortium. The MRDF bibliographer is the Official Representative of UCSD to the Consortium.

B. Physical acquisitions of MRDFs and documentation:

Currently, magnetic tape is the primary medium for acquisitions of machine-readable information. In addition, the library acquires some MRDFs in other formats such as personal computer floppy disks, CD-ROM, and by electronic transfer (e.g., electronic mail and downloading from bulletin boards).

The Library acquires the necessary datafile documentation (e.g., codebooks, user guides, technical documentation) in hardcopy, machine-readable, or micro-form format. The particular format chosen depends on availability (some documentation is available in only one format), ease of use (large documentation may sometimes be more easily searched than unwieldy paper copy or microfiche), and size and cost (some documentation is so voluminous that printed copies are very expensive).

C. Storage of MRDFs and documentation:

Working copies of most datafiles are kept on one-half inch computer tape. Several files are stored on magnetic disk on the machine webber as part of the SSDB. In addition, some files are kept in their original non-tape format (e.g., CD-ROM, floppy disk).

Working copies of tapes are stored in the Instructional Computing Center. Backup and original copies are stored at the Library Annex. Floppy disks and CD-ROMs are stored in the Central Library.

As of the Fall of 1990, the Library will begin to move working copies of datafiles to re-writable magneto-optical disk storage.

Paper copy and microform documentation are kept in the Library and integrated into the collections there. Machine-readable documentation is stored on tape or disk with the datafile documented.

IV. Other available resources

The Inter-University Consortium for Political and Social Research is the primary archive for machine-readable social science data in the United States.

The Institute for Social Science Research at UCLA and the State Data Archives at UCB maintain large collections of social science data in machine readable form.