

CORNELL UNIVERSITY  
Cornell Institute for Social and Economic  
Research Policy

## CISER Data Archive Collection Policy

POLICY  
Volume: DA

Responsible Executive:  
CISER Data Librarian

Responsible Office:  
Cornell Institute for  
Social and Economic  
Research

Originally Issued:  
Revised: 4/3/14

### POLICY STATEMENT

---

CISER Data Archive is a centralized source for numeric datasets: their acquisition, storage, maintenance, use, and preservation. It is integral to CISER's mission which is to anticipate and support the evolving computational and data needs of Cornell social scientists and economists throughout the entire research process and data lifecycle.

### GUIDELINES

---

CISER houses an extensive collection of public and restricted digital research data files in the social sciences with particular emphasis on data that matches the interests of Cornell researchers. In some cases data held in the Data Archive are unique copies and cannot be found elsewhere.

Core clientele are faculty, graduate students, and research staff in schools and colleges that support CISER, namely:

- College of Agriculture and Life Sciences
- College of Arts and Sciences
- College of Human Ecology
- Johnson Graduate School of Management
- School of Hotel Administration
- School of Industrial and Labor Relations

CISER also support users from other Cornell schools whose researchers have a need for social science data.

CISER also hosts Cornell Restricted Access Data Center (CRADC), the University custodian of restricted access data sets. CRADC staff work closely with data suppliers to provide a secure environment for restricted-use datasets.

## SUBJECTS COVERED

---

CISER collects and maintains digital research data files in the social sciences, with emphasis on demography, economics and labor, political and social behavior, family life, and health. The collection includes, but not exclusively so, federal or state censuses, files based on administrative records, public opinion surveys, economic and social data from national and international organizations, and studies compiled by Cornell researchers.

The Data Archive acquires or accepts data for any geographic area. The collection predominantly deals with the United States, European countries and international data produced by intergovernmental organizations. Emphasis is also placed on data related to New York State. Although it is often difficult to obtain data from some developed nations and most developing states, the CISER Data Archive works with researchers to meet their needs.

## PURCHASE OF DATA DATASET APPRAISAL AND ACQUISITION

---

Appraisal is accomplished by CISER staff in conjunction with:

- recommendations from faculty
- an evaluation of the quality of the data and the reliability of the distributor
- expected future use by a broad constituency of social science users.

Data acquisition is primarily demand driven. The Data Archive will attempt to acquire any set of data required by faculty members in accordance with organizational policies regarding cost, quality, restrictions, and expected future use by a broad constituency of social science and economics users.

Using the same criteria, data are also acquired for students of those faculty who are engaged in substantive social science or economic research. Pro-active collection development is undertaken in anticipation of demand.

The Data Archive will maintain data series identified as core to the data collection.

Due to contractual agreements between Cornell University and the Inter- university Consortium for Political and Social Research (ICPSR) members of the Cornell Community are entitled to

obtain any of the data offerings of the Consortium. CISER Data Archive actually serves any and all members of the community in terms of data acquisitions from the Consortium, regardless of subject area.

## PURCHASE OF DATA

---

When data purchases are requested by researchers, the Data Archive may support the entire purchase or a part. Alternatively, another department or individual may acquire/purchase the data and deposit the file at the archive. Assistance in the processing of the request is available, with the archive acting as an interface between the data source and the requester.

When a data request is initiated by an individual, the requester will be asked to provide the staff with a description of the data, written justification for the purchase of the file, and a cost estimate for data acquisition. It may be recommended that the requester go directly to another funding source, such as his own department, another agency, or cooperate in pooling resources.

The Data Archive works with Library Collection Development staff, faculty, and departments to secure full or matching funding, especially in cases where a dataset has a potential audience representing more than one academic department. The Data Archive also collaborates with Cornell libraries and other information services at Cornell to assure that collection content and access are not duplicated, so long as CISER clients can use data and material from those units with reasonable effort. When acquiring material, the Archive must consider not only content but format and delivery criteria to fulfill its mission and meet the needs of its clientele.

## FILE FORMATS

---

CISER offers its depositors a list of preferred and acceptable file formats. The formats are commonly used within the social science and economics domain, have open specifications, and are independent of specific software, developers or suppliers. CISER will however accept data regardless of physical format as long as they are convertible to supported and accessible file formats suited for long-time preservation for use by the entire Cornell community.

Where possible CISER will normalize data in proprietary formats into accompanying raw ASCII or Unicode

Staff will also work with users or other campus information services to migrate legacy formats, including data on obsolete physical media, to insure continued use of unique studies.

## CONFIDENTIALITY

---

CISER Data Archive will not accept data which contains personal identifiers, except in such cases where these data are part of the public record. Datasets held in the archive are primarily public-use versions. For restricted access and limited use data products CRADC provides secure access.

## DOCUMENTATION

---

Where possible data studies are accompanied by comprehensive machine-readable documentation: codebooks, file layout maps, technical notes, questionnaires, reports, and errata in open and accessible formats. Non-digital documentation is available when machine-readable documentation is not.

In cases where documentation is insufficient CISER works with data producers to ensure that data files are useable and understandable by generating additional contextual information.

## DATA QUALITY

---

The Data Archive reserves the right to reject datasets deemed to be inadequately documented, are potentially disclosive, are acquired or generated illegally, or are suspected or known to contain inaccuracies.

## RESPONSIBILITIES

---

The following are major responsibilities each party has in connection with this policy.

CISER Director / CISER Data Librarian	Interpret this policy and provide clarification and education and implement operational and business processes to facilitate compliance.
CRADC Director of Information Technology	Implement operational, physical, and technical equipment and tools to facilitate compliance.

Any employee found to have violated this policy may be subject to disciplinary action, up to and including termination of employment.

## CONTACTS

---

If you have questions about specific issues regarding this Data Collection Policy, call the following offices:

William Block	<a href="mailto:block@cornell.edu">block@cornell.edu</a>	CISER Director	607-255-9026
Data Librarian	<a href="mailto:ciser@cornell.edu">ciser@cornell.edu</a>	Data Librarian	607-255-4801

---