CORNELL UNIVERSITY
Cornell Institute for Social and Economic Research Policy

POLICY
Volume: DA

Responsible Executive: CISER Data Librarian

Responsible Office: Cornell Institute for Social and Economic Research

Originally Issued:
Revised: 4/3/14

# CISER Data Archive Preservation and Storage Policy

## POLICY STATEMENT

The data preservation function is integrated into the operations and planning of CISER and throughout the management stages of the research data lifecycle in order to support Social Science and Economic research at Cornell University.

## REASON FOR POLICY

The fundamental purpose of CISER's Data Archive is to select, preserve and make available for use primary and secondary data, documentation and metadata, in discipline recognized digital formats that remain suitable for research in perpetuity. The data preservation and storage policy is guided by a variety of community-driven standards, (e.g. Open Archival Information Systems (OAIS) reference model, Trusted Repositories Audit and Certification (TRAC), Data Seal of Approval (DSA), Data Documentation Initiative (DDI)), that represent an international body of knowledge and expertise pertaining to various issues within digital preservation.

## POLICY GUIDELINES

These guidelines address the effective implementation of procedures for the preservation of CISER's digital collections within the context of the *CISER Data Archive Collection Policy*. CISER reserves the right to review the scholarly and historical value and user accessibility into the data preservation characteristics.

**Data Integrity**

Upon receipt of new digital content, the Archive staff process the data and documentation, assess that proper confidentiality concerns are addressed, in collaboration with the data producer fix errors if necessary, convert data formats, and run a checksum. The metadata pertaining to each data file is stored in a SQL database. (A backup of the SQL database is taken every evening and is retained for a finite period.) Provenance notes are maintained, which relate back to the original deposited version, as part of the metadata for any alterations made in the preservation and dissemination versions.

To ensure that the digital content remains identical and accessible, scripts are run on a scheduled basis to verify checksum, permissions, and record counts. The results are compared to the metadata, held within the SQL database, to validate data integrity.

If degradation of any digital content is detected, CISER would endeavor to re-instate the original version from a backup copy. After data retrieval scripts are then run to ensure data integrity has not been compromised.

**Data Normalization**

Evaluation of new content types and software/format obsolescence is an ongoing process. It is expected that normalizing the Data Archive collection by migrating to updated content types when new formats become widely available occur seamlessly. When new formats are created from data files either through migration into new file formats or through creating new file formats for dissemination, the old files are retained alongside. Version control is stored as part of the metadata, as referenced in the *CISER Data Archive Versioning Policy*.

**Management of Storage Infrastructure**

The preservation of the Data Archive is dependent upon CISER's storage infrastructure. Thus, management of the storage infrastructure is designed to accommodate scalability, reliability, and sustainability, in accordance with quality control specifications and security regulations. In lite of increasing user demand and changing technologies, CISER staff routinely monitors technical developments and evaluates potential archival solutions that will both streamline and enhance CISER data preservation practices.

Adequate storage capacity for all Data Archive holdings is maintained. In addition, unlimited capacity from external media is available. The disk storage maintains a RAID 6 configuration and all infrastructures are protected by uninterrupted power supplies (UPS).

The Data Archive holdings are maintained in both a compressed and uncompressed state. The compressed data are readily available for download, while the uncompressed are available to utilize on CISER's computing servers using a multitude of available statistical applications. Both versions (compressed and uncompressed) are backed up on a daily basis via the University's offering of EZ-backup, which also provides off-site storage. EZ-backup makes use of IBM's Tivoli Storage Manager.

**Security**
CISER is committed to taking all necessary precautions to ensure the physical safety and security of the Data Archive holdings that it preserves. The storage infrastructure is housed in the University data center. The data center features uninterrupted power supplies (UPS), fire prevention and protection system, physical intruder prevention and detection systems and environmental control systems. In addition, the server racks that house the CISER's disk storage are equipped with unique keys. Only CISER's system administration staff have access to the server racks.

## Entities Affected by this Policy

CISER Director, CISER Data Librarian, CISER Information Technology staff, internal and external Data Archive Users, and Office of Sponsored Programs.

## Who Should Read this Policy

CISER Director, CISER Data Librarian, CISER Information Technology staff, Office of Sponsored Programs, and Data Archive Users.

## Related Documents

CISER Data Archive Collection Policy
http://ciser.cornell.edu/pub/policies/CISER_Data_Collection_Policy.pdf

CISER Data Archive Versioning Policy
http://ciser.cornell.edu/pub/policies/CISER_Data_Versioning_Policy.pdf

Cornell University Policy 4.12, Data Stewardship and Custodianship
https://www.dfa.cornell.edu/tools-library/policies/data-stewardship-and-custodianship

Cornell University Policy 5.1, Responsible Use of Information Technology Resources
https://www.dfa.cornell.edu/tools-library/policies/responsible-use-information-technology-resources

Cornell University Policy 5.4.1, Security of Information Technology Resources
https://www.dfa.cornell.edu/tools-library/policies/security-information-technology-resources

Cornell University Policy 5.10, Information Security
https://www.dfa.cornell.edu/tools-library/policies/information-security

Cornell University Policy 4.6, Standards of Ethical Policy
https://www.dfa.cornell.edu/tools-library/policies/standards-ethical-conduct

## CONTACTS

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

| | | | |
|---|---|---|---|
| William Block | block@cornell.edu | CISER Director | 607-255-4801 |
| Data Librarian | ciser@cornell.edu | CISER Data Librarian | 607-255-4801 |
| Janet Heslop | jheslop@cornell.edu | CISER Director of Information Technology | 607-255-8531 |

## 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. |
| CISER Director of Information Technology | Implement operational, physical, and technical equipment and tools to facilitate compliance. |
| CISER Data Librarian | Responsible for performing tasks in accordance with established policy guidelines. |
| CISER Affiliated Staff | Responsible for knowing and assisting with tasks as needed related to this policy. |

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