Implementation of the Data Seal of Approval

The Data Seal of Approval board hereby confirms that the Trusted Digital repository CISER Data Archive complies with the guidelines version 2 of 2013 set by the Data Seal of Approval Board.
The afore-mentioned repository has therefore acquired the Data Seal of Approval of 2013 on May 19, 2014.

The Trusted Digital repository is allowed to place an image of the Data Seal of Approval logo corresponding to the guidelines version date on their website. This image must link to this file which is hosted on the Data Seal of Approval website.

Yours sincerely,

The Data Seal of Approval Board

Assessment Information

Guidelines Version: 2 | July 19, 2013
Guidelines Information Booklet: DSA-booklet_2_July2013.pdf
All Guidelines Documentation: Documentation

Repository: CISER Data Archive
Seal Acquiry Date: May. 19, 2014

For the latest version of the awarded DSA   http://assessment.datasealofapproval.org/seals/
for this repository please visit our website:

Previously Acquired Seals: None

This repository is owned by: **Cornell Institute for Social and Economic Research (CISER)**
  *

USA

T 607 255 4801
E ciser@cornell.edu
W http://ciser.cornell.edu/

Assessment

0. Repository Context

Applicant Entry

Self-assessment statement:

Assessment

0. Repository Context

Applicant Entry

Self-assessment statement:

The Cornell Institute for Social and Economic Research (CISER), founded in 1981, is home to one of the oldest university-based social science data archives in the United States. CISER houses an extensive collection of public and restricted numeric data files in the social sciences with particular emphasis on studies that match the interests of Cornell researchers: demography, economics and labor, political and social behavior, family life, and health.

CISER's mission is to anticipate and support the evolving computational and data needs of Cornell social scientists and economists throughout the entire research process and data life cycle. Data archive functions include making data available to the broadest audience permissible via green/yellow/red light access levels; providing a secure, safe research computing environment to facilitate data access and use for researchers; and data consulting support from staff experienced in using social science data, in order to maximize the benefits of the data archive and research computing facilities, including availability of our significant depth of expertise in restricted data access management.

We do not outsource any of our functions. We manage all aspects of support for the researcher throughout the data life cycle.

Links to supporting documentation:

CISER: http://ciser.cornell.edu [accessed 4/4/14]

CISER Data Archive Policies: http://ciser.cornell.edu/pub/policies/CISER_Policies.shtm [accessed 4/4/14]

Cornell University Research Data Management Services Group (RDMSG): http://data.research.cornell.edu [accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data, and compliance with disciplinary and ethical norms.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data, and compliance with disciplinary and ethical norms.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

CISER works with a range of national and international organizations as well as individual researchers to receive their data files. The majority are public use files that are placed in the CISER Data Archive for widespread access. Some are confidential and housed within the Cornell Restricted Access Data Center (CRADC). CISER employs the highest standard of ingest processing to ensure the quality and integrity of datasets.

Upon receipt of new data, Data Archive staff verify the integrity of the digital data by running multiple scripts to determine number of records, variables, bytes, file names, checksum, etc. Errors are corrected if necessary, data are formatted to meet discipline standards, and confidentiality concerns are addressed. Digital content metadata is then stored in a SQL database and is searchable at study level and file name. The SQL database is backed up nightly via a scheduled SQL job. Regular scripts are re-run to ensure the digital collection remains identical (i.e. checksum, bytes, number of records) and accessible according to granted permissions. See CISER Data Preservation and Storage Policy for further details.

CISER works with the data providers to resolve any missing information, inconsistencies, and confidentiality issues that may be found during this stage. CISER checks the documentation provided by the data provider for completeness. If incomplete, CISER works with the data provider to gather more information/documentation. Hardcopies of the documentation are converted into electronic form using the PDF/A format for archival and downloading purposes.

Metadata creation continues after the initial processing as this is a process that is undertaken across the data life cycle (i.e., from data conceptualization to collection, processing, distribution, discovery, analysis, repurposing, and archiving). It is highly likely that additional user information will be provided, such as a Readme file or other documents that detail the changes that were made to the original data and/or other instructions for using the collection.

For restricted access datasets, secure data providers ship data on physical media (disk, portable drive) using a delivery service that enables tracking of the package, or transmit the files electronically using a secure service, such as Cornell DropBox. Files, whether shipped on media or transmitted electronically, are encrypted by a process that meets or exceeds specified security standards. Upon receipt, data are transferred to a secure file server with original files being securely stored on different media for safe keeping. CISER works with data providers to implement security plans meeting provider requirements. Part of this process is to review for disclosure risks.

Links to supporting documentation:

CISER Data Archive Preservation and Storage Policy:http://ciser.cornell.edu/pub/policies/CISER_Data_Preservation_and_Storage_Policy.pdf[accessed 4/4/14]

CISER Data Archive Collection Policy: http://ciser.cornell.edu/pub/policies/CISER_Data_Collection_Policy.pdf[accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

2. The data producer provides the data in formats recommended by the data repository.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

2. The data producer provides the research data in formats recommended by the data repository

Minimum Required Statement of Compliance

3. In progress: We are in the implementation phase.

Applicant Entry:

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

In order to guarantee the use of data both now and in the future it is important that datasets are archived in supported and accessible formats. CISER, therefore, offers its depositors a list of preferred and acceptable formats

that it considers best suited for long-time preservation and accessibility. The file formats are commonly used within the social science and economics domain, have open specifications, and are independent of specific software, developer or supplier. CISER is willing to accept research data in other formats, if they are convertible to open and available file formats. Where possible CISER will normalize data in proprietary formats into accompanying raw ASCII or Unicode.

During the ingest process a detailed standard routine is followed to check validity and quality of data files and asks that depositors whose datasets contain file types different from listed formats contact the data archive. CISER staff check submitted datasets for their file formats and contact the depositor, if necessary.

CISER staff expect that the list of preferred and acceptable formats will change over time as new formats are developed and others fall into disuse.

Links to supporting documentation:

CISER Data Archive Collection Policy: http://ciser.cornell.edu/pub/policies/CISER_Data_Collection_Policy.pdf [accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

3. The data producer provides the data together with the metadata requested by the data repository.

Minimum Required Statement of Compliance:

    4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

    4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

    3. The data producer provides the research data together with the metadata requested by the data repository.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Data producers seeking to deposit data in the CISER Data Archive must provide metadata in compliance with domain standards. Where possible data studies should be accompanied by comprehensive machine-readable documentation: codebooks, file layout maps, technical notes, questionnaires, reports, and errata in open and accessible formats.

This can be facilitated through CISER's data deposit process, which currently makes use of tools that generate DDI XML from commonly-utilized statistical software packages (such as SAS, SPSS, Stata), and will soon include a web interface for increased efficiency. CISER provides depositors with assistance to ensure that the metadata produced will be sufficient for ingest. The metadata provided, along with basic contact information for a given data producer, is collected at the time of request, and stored in a relational database. If more information is needed, the data producer is contacted for follow up. The relational database drives the online search interface of the CISER Data Archive, where study and file-level information can be accessed.

Links to supporting documentation:

CISER Data Archive online catalog: http://ciser.cornell.edu/ASPs/search.asp[accessed 4/4/14]

CISER Data Archive Collection Policy: http://ciser.cornell.edu/pub/policies/CISER_Data_Collection_Policy.pdf [accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

4. The data repository has an explicit mission in the area of digital archiving and promulgates it.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

4. The data repository has an explicit mission in the area of digital archiving and promulgates it.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

As one of the oldest university-based social science data archives in the United States, CISER has demonstrated its commitment to the long-term preservation and access of data for scientific research.

CISER's mission to anticipate and support the evolving computational and data needs of Cornell social scientists and economists throughout the entire research process and data life cycle is integrated into organizational policy, procedure, and practice. Facilities include: a state-of-the-art computing cluster of multi-processor Windows servers; expansive disk storage and daily backups; access to statistical software packages (e.g. SAS, SPSS, Stata, Gauss, R, Matlab, Stat Transfer); and a separate, secure computing environment to support use of confidential datasets.

The CISER Data Archive holds an extensive collection of numeric files in the social sciences, with emphasis on demography, economics and labor, political and social behavior, family life, and health. It provides consulting services to identify, obtain, and use datasets, with fully trained staff whom work with data providers to ensure that data and accompanying documentation comply with CISER standards and policies. CISER also provides access to a substantial web-based library of sample programs and instructional material on using the CISER research servers.

CISER staff actively promulgate our mission by being extensively involved in the international social science data community and profession, including membership, committee work, and holding leadership roles in organizations such as the International Association for Social Science Information Service and Technology (IASSIST), the Association of Public Data Users (APDU), and the Data Documentation Initiative (DDI). CISER also promotes its mission through publications, attendance at conferences, and through supporting a rich array of professional development activities for CISER staff.

CISER is also home to the Cornell Restricted Access Data Center (CRADC) which provides a secure environment for remote access to restricted-use datasets and the New York Census Research Data Center (NYCRDC) which provides academic researchers with access to selected Census confidential microdata in physically secure facilities.

Links to supporting documentation:

CISER website: http://ciser.cornell.edu/[accessed 4/4/14]

CISER Data Archive Mission Statement: http://ciser.cornell.edu/pub/policies/CISER_Mission_Statement.pdf [accessed 4/4/14]

CISER Data Archive Preservation and Storage Policy: http://ciser.cornell.edu/pub/policies/CISER_Data_Preservation_and_Storage_Policy.pdf [accessed 4/4/14]

CISER Data Archive Collection Policy: http://ciser.cornell.edu/pub/policies/CISER_Data_Collection_Policy.pdf [accessed 4/4/14]

CISER Terms of Use: http://ciser.cornell.edu/pub/policies/CISER_Terms_of_Use.pdf [accessed 4/4/14]

CRADC: http://ciser.cornell.edu/CRADC/What_is_CRADC.shtml [accessed 4/4/14]

NYCRDC: http://ciser.cornell.edu/NYCRDC/home.shtml [accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

CISER is legally considered part of Cornell University and is housed within the Office of the Vice Provost for Research. All CISER Data Archive users must agree to a Terms of Use Policy prior to gaining access to data held in the archive. The policy explains that: users are responsible for complying with all applicable federal, state and local laws; they must abide by Cornell University policies; and agree to adhere to data provider licensing requirements.

The dissemination of data from the CISER Data Archive is built upon a "green-yellow-red" light system. The files that are publicly available are declared with a "green light", those with a "yellow light" are limited to Cornell affiliated researchers only, while those classified with a "red light" are restricted and require permission prior to use as stipulated by respective data providers.

Other legal contracts and regulations that CISER handles are primarily related to Data Provider Agreements (DPA) for restricted-use datasets within the Cornell Restricted Access Data Center (CRADC). The DPA stipulates use, dissemination, and backup specifications of restricted-use data. All DPAs are evaluated by the Office of Sponsored Programs (OSP) and the Institutional Research Board (IRB) for terms and regulations governing the protection of human subjects. In addition, the DPA includes information on penalties for noncompliance. OSP negotiates these agreements if necessary, and signs the DPA on behalf of Cornell University.

Data held by CRADC are categorized restricted-access and are only available via a legally signed contract with the DPA. An entirely separate computing domain and servers are built specifically for this function, as documented in an Information System Security Plan (confidential document) per NIST 800-18 guidelines. Data Custodians are trained in handling restricted-use data and must comply by renewing the training on a regular basis.

With respect to compliance with national laws under which CISER operates, in the United States there are several statutes and codes related to the privacy and protection of research participants. Of particular note is the federal regulation on Protection of Human Subjects (45 CFR 46). Institutions bear the responsibility for compliance with 45 CFR 46. Every university must file an "assurance of compliance" with the Office of Research Integrity Assurance which includes "a statement of ethical principles to be followed in protecting human subjects of research." University Institutional Review Boards (IRBs) review research to address these issues. Other relevant U.S. laws include the Family Educational Rights and Privacy Act (FERPA), the Health Insurance Portability and Accountability Act (HIPAA), the Confidential Information Protection and Statistical Efficiency Act (CIPSEA), and Title 13 US Code – Protection of Confidential Information.

Links to supporting documentation:

CRADC: http://ciser.cornell.edu/CRADC/What_is_CRADC.shtml[accessed 4/4/14]

CRADC - Steps to Acquire and Use Restricted Data:
http://ciser.cornell.edu/CRADC/ObtainingRestrictedData.shtm[accessed 4/4/14]

Cornell University Institutional Review Board for Human Participants: http://www.irb.cornell.edu/[accessed 4/4/14]

Cornell University Office for Sponsored Programs: http://www.osp.cornell.edu/[accessed 4/4/14]

CISER Terms of Use: http://ciser.cornell.edu/pub/policies/CISER_Terms_of_Use.pdf[accessed 4/4/14]

CISER Data Archive Security Policy:
http://ciser.cornell.edu/pub/policies/CISER_Data_Security_Policy.pdf[accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

6. The data repository applies documented processes and procedures for managing data storage.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

6. The data repository applies documented processes and procedures for managing data storage.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The CISER Data Archive is stored on network attached storage (NAS) in both compressed and uncompressed format in Cornell University's Data Center. The compressed data is for public download access via the CISER data catalog. The uncompressed data is accessible on the CISER computing servers. The NAS disk runs RAID 6 and has manufacture call-home features enabled for expedited servicing.

The dissemination of data from the CISER Data Archive is built upon a "green-yellow-red" light system. The files that are publically available are declared with a "green light", those with a "yellow light" are limited to Cornell affiliated researchers only, while those classified with a "red light" are restricted and require permission prior to use as stipulated by respective data providers.

Backups are performed daily using Tivoli Storage Manager (TSM) offered as a service named EZ-Backup from Cornell's Central IT Office. EZ-Backup provides an offsite storage facility in New York City. In addition, three copies of changed files are kept in the backup database at all times. Deleted files remain available for 180-days. Data recovery can be accomplished by the CISER Systems Administrative staff or the EZ-backup Team. In the event of disaster, the EZ-backup Team would be the primary contact for restoring the CISER Data Archive.

In accordance with Cornell University Emergency Response plans CISER is in the process of developing a Continuity of Operations (COOP) plan which details processes and procedures to deal with data and service recovery in the event of disaster. The COOP plan will addresses: Prevention and risk mitigation (incl. risk assessment); Preparedness (incl. emergency exercise development); Response (incl. emergency operations plan and support functions); Recovery (incl. University recovery plan, CISER essential functions)

For restricted-access data CRADC staff manage the restricted data in a manner conforming to the data providers' terms and conditions, including the CRADC security plan. The restricted access files are not backed-up and there is no archiving. Only working files containing no restricted data can be restored to the CRADC secure file servers in case of lost or corrupted files. The metadata is not searchable but rather is stored with the restricted data files and is only accessible by authorized and authenticated users.

Restricted access files are kept on the CRADC secure file server in a remote site. The original restricted access data files supplied by the data providers are stored on physical media in a fireproof safe in the CISER building. These files can be securely transferred to the CRADC server in case the online files need to be restored. The working files—which do not contain copies of the restricted access data—are backed up over a secure black fiber connection to a disk backup system in the CISER machine room. The backup system meets FIPS 140-2 compliancy regarding security. The data use agreements with data providers typically require that at the end of the project period—anywhere from 1 to 5 years—the original media be returned or destroyed and that all copies of the data be destroyed.

Links to supporting documentation:

CISER Data Archive Preservation and Storage Policy:
http://ciser.cornell.edu/pub/policies/CISER_Data_Preservation_and_Storage_Policy.pdf [accessed 4/4/14]

CISER Data Archive Security Policy: http://ciser.cornell.edu/pub/policies/CISER_Data_Security_Policy.pdf [accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

7. The data repository has a plan for long-term preservation of its digital assets.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

7. The data repository has a plan for long-term preservation of its digital assets.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4 Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The CISER Data Preservation and Storage Policy documents the main theoretical and practical steps for providing long-term preservation of digital research data. Data preservation is integrated into archival operations and

planning within CISER as part of the research data lifecycle.

CISER ensures the integrity, completeness, and authenticity of data submitted to the Data Archive during the ingest process as outlined in our Data Collection Policy. During the ingest process, non-supported file formats are converted to specified formats that support long-term preservation.

CISER routinely monitors technical developments (standards, software, tools, and platforms) and evaluates potential archival solutions that will both streamline and enhance CISER data preservation and archival practices.

Links to supporting documentation:

CISER Data Archive Preservation and Storage Policy:
http://ciser.cornell.edu/pub/policies/CISER_Data_Preservation_and_Storage_Policy.pdf [accessed 4/4/14]

CISER Data Archive Collection Policy: http://ciser.cornell.edu/pub/policies/CISER_Data_Collection_Policy.pdf
 [accessed 4/4/14]

CISER Data Archive Security Policy: http://ciser.cornell.edu/pub/policies/CISER_Data_Security_Policy.pdf
 [accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

8. Archiving takes place according to explicit work flows across the data life cycle.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

8. Archiving takes place according to explicit flows across the data lifecycle

Minimum Required Statement of Compliance:

3. In progress. We are in the implementation phase.

Applicant Entry:

Statement of Compliance

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

CISER procedures follow the data life cycle and adhere to predetermined criteria that apply at each stage. These include:

- data management planning support for grant funded research;

- data processing procedures (data manipulation and reformatting; integration and/or harmonization of data series; simulated and synthetic data for training on confidential data sets):

- data documentation (development of comprehensive metadata):

- data discovery and re-use via the Data Archive catalog

- data preservation (data integrity, normalization, storage infrastructures)

The CISER Director was a founding member and CISER staff are key participants in Cornell University's Research Data Management Service Group (RDMSG). The RDMSG provides timely and professional assistance for the creation and implementation of data management plans and helps researchers find specialized data management services they require at any stage of the research process (including initial exploration, data gathering, analysis and description, long term preservation and access).

CISER staff who manage data have a set of internal guidelines that they adhere to and they document ingest processes and data transformations. Other processes such as long-term preservation (e.g. normalization, version control, sustainability) are detailed in the CISER Data Preservation and Storage Policy

The CISER Data Collection Policy details criteria and information regarding the selection of data for archiving. Storage is managed according to strict criteria regarding media, redundancy, and is detailed in the CISER Data Preservation and Storage and CISER Data Security Policies. CISER staff are developing data evaluation and appraisal templates for evaluating the new content types and software/format obsolescence.

Links to supporting documentation:

CISER Data Archive online catalog: http://ciser.cornell.edu/ASPs/search.asp[accessed 4/4/14]

RDMSG:
https://confluence.cornell.edu/display/rdmsgweb/Home;jsessionid=584576AE814BC4B4F83360DFF36ED6FF[accessed 4/4/14]

CISER Data Archive Preservation and Storage Policy:
http://ciser.cornell.edu/pub/policies/CISER_Data_Preservation_and_Storage_Policy.pdf [accessed 4/4/14]

CISER Data Archive Collection Policy:
http://ciser.cornell.edu/pub/policies/CISER_Data_Collection_Policy.pdf[accessed 4/4/14]

CISER Data Archive Security Policy:
http://ciser.cornell.edu/pub/policies/CISER_Data_Security_Policy.pdf[accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

9. The Data repository assumes responsibility from the data producers for access and availability of the digital objects

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

4. Implemented: This guidelines has been fully implemented for the needs of our repository

Self-assessment statement:

For public-use datasets CISER complies on a case-by case basis with data producer terms and conditions through data producer agreements signed by the Data Librarian. In addition CISER has annual memberships with ICPSR and the Roper Center for Public Opinion Research and other organizations. Such agreements make available and accessible datasets from these data producers for Cornell students, faculty and researchers.

CISER staff strives to meet a high standard of ingest processing to improve the quality of datasets. We work closely with the data providers to resolve any missing information or documentation, inconsistencies, and confidentiality issues that may be found during this stage. For further information refer to the CISER Data Collection Policy.

CISER's Cornell Restricted Access Data Center (CRADC) manages user access in conformance with legal contracts/regulations primarily related to Data Provider Agreements (DPA) for restricted-use datasets. The DPA stipulates use, dissemination, and backup specifications of the data. All DPAs are evaluated by the Office of Sponsored Programs (OSP) and the Institutional Research Board (IRB) for terms and conditions governing the protection of human subjects. In addition, the DPA includes information on penalties for noncompliance. OSP negotiates, if necessary, and signs the DPA on behalf of Cornell University.

An entirely separate domain and servers are built specifically for restricted-use datasets, as documented in an Information System Security Plan (confidential document) per NIST 800-18 guidelines. CRADC staff are trained in handling restricted-use data and must comply by renewing the training on a regular basis.

In accordance with Cornell University emergency response plans CISER is in the process of developing a Continuity of Operations (COOP) plan which details process and procedures to deal with data and service recovery in the event of disaster. See Guideline 5.

Links to supporting documentation:

CISER System Usage Agreement - http://ciser.cornell.edu/computing/manual/useAgreement.shtm [accessed 4/4/14]

CISER System Usage Policies - http://ciser.cornell.edu/computing/manual/policies.shtm [accessed 4/4/14]

CISER Data Archive Use Policies - http://ciser.cornell.edu/info/policy.shtml [accessed 4/4/14]

Cornell University Policy Office (Policy Volume V. Information Technologies - http://www.dfa.cornell.edu/treasurer/policyoffice/policies/volumes/informationtech/index.cfm[accessed 4/4/14]

Cornell University IT Policy and Law - http://www.it.cornell.edu/policies/university/privacy/abuse/index.cfm [accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

10. The data repository enables the users to discover and use the data and refer to them in a persistent way.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

10. The data repository enables the users to utilize the research data and refer to them

Minimum Required Statement of Compliance:

3. In Progress. We are in the implementation phase

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The CISER Data Archive provides access to social science, economic, and health research data and documentation in formats required and used by the Cornell research community such as SPSS, SAS, Stata, PDF, and raw ASCII and Unicode.

The CISER Data Archive online catalog offers robust search facilities to enable discovery of and access to both public-use and restricted-use data files held on the CISER file server with legacy data held on CDROM/DVD (scheduled to be moved on to the file server in 2014). Users are also able to download codebooks and other documentation materials through the catalog. The Data Archive collection is preserved by migrating the collection to new versions or when new formats become widely available. Users can search for data by title, producer, principal investigator in addition to conducting free text searching with truncation. The Catalog can also be browsed by subject area.

Data files, documentation, and ancillary files are housed on the CISER research computing servers which allow CISER computing account holders to prepare, analyze and manage data using statistical software packages (e.g. Atlas.ti, Gauss, Mathematica, Matlab, R, SAS, SPSS, Stata) For complete list see http://ciser.cornell.edu/computing/software.shtml [accessed 4/4/14].

All data studies maintained by the CISER Data Archive are assigned a locally-generated unique identifier, and will be assigned a study-level DOI using the California Digital Library's EZID service (except those prohibited by the data provider in the case of restricted access files). Documentation provided with each study includes a standard format study-level citation. Some data collections have a different access condition level which prevent data consumers (users) who do not meet the relevant criteria from accessing them.

Restricted-access datasets residing on CRADC are not catalogued, however, they can be accessed remotely and securely 24/7 by approved researchers based on the limits set in their signed data use agreement. Limited-use licensed data can either be accessed via CRADC or, if approved by the provider, delivered to researchers in CD or DVD format. After approval, the CRADC Data Custodian will customize secure computing accounts for each member of the project team to allow access to the data.

Links to supporting documentation:

EZID - http://ezid.cdlib.org/ [accessed 4/4/14]

CISER Data Archive online catalog: http://ciser.cornell.edu/ASPs/search.asp [accessed 4/4/14]

CISER Research Computing: http://ciser.cornell.edu/pub/Resources.shtml [accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

11. The data repository ensures the integrity of the digital objects and the metadata.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

11. The data repository ensures the integrity of the digital objects and the metadata.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

CISER accepts responsibility for preserving and making available digital content, associated documentation and other metadata provided by depositors in accordance with the CISER Data Archive Collection Policy.

A checksum is generated by using a the MD5 File Hasher utility for every data file added to the CISER Data Archive to ensure integrity of the digital file both now and into the future. Data versioning criteria are consistently applied to changes in data files and data documentation (such as correction for error, documentation amendments, additional variables, changes in access conditions, format changes) for inclusion in the CISER Data Archive. Once deposited, files in datasets are never changed and only minor changes to the metadata are allowed. Changes to the data themselves are issued as a new version of the dataset (which will be issued a new study-level persistent identifier (DOI) using the California Digital Library's EZID service). This often involves working closely with the data producer. Scripts are run on a regular basis:

- New Technology File System (NTFS) file permissions are checked to:

1. Verify that restricted files have restricted permission settings on the file server.

2. List which researchers have access to restricted files.

- Path/filename comparisons are run to ensure the archive path and filename match the path and filename metadata in the catalog.

- MD5 checksum validations are run to report all files which have been added, deleted, or modified since the previous validation.

Files which are available for download are kept in publicly accessible folders or in folders that can only be accessed with Cornell University credentials to ensure that studies are available to suitably approved users.

Links to supporting documentation:

CISER Data Archive Collection Policy: http://ciser.cornell.edu/pub/policies/CISER_Data_Collection_Policy.pdf
 [accessed 4/4/14]

CISER Data Archive Versioning Policy: http://ciser.cornell.edu/pub/policies/CISER_Data_Versioning_Policy.pdf
[accessed 4/4/14]

CISER Data Archive Preservation and Storage Policy:
http://ciser.cornell.edu/pub/policies/CISER_Data_Preservation_and_Storage_Policy.pdf [accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

12. The data repository ensures the authenticity of the digital objects and the metadata.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

12. The data repository ensures the authenticity of the digital objects and the metadata.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

CISER is currently working with the California Digital Library EZID service to assign a DOI at the study level upon ingest. Subsequent versions of data can be ingested, and assigned separate DOIs. The W3C PROV standard

will be employed to encode and store provenance of a given record. This also provides a means to backtrack up the version chain. If changes need to be made to the data that do not warrant a version change, arrangements can be made through the CISER help desk. New versions can be submitted through our online deposit form (in development) with assistance from CISER staff according to CISER's Data Collection Policy. Depositors sign in through Cornell Kerberos authentication. We also have plans to implement Shibboleth for federated identity management. Data is backed up by Cornell Information Technology EZ-Backup service. Metadata is backed up through SQL Server jobs nightly and is held for two-weeks before deleting. On a semi-annual basis the entire SQL database is backed-up and stored in a permanent location, which includes an off-site replicate.

Also see Guideline 3.

Links to supporting documentation:

CISER Data Archive Preservation and Storage Policy:
http://ciser.cornell.edu/pub/policies/CISER_Data_Preservation_and_Storage_Policy.pdf [accessed 4/4/14]

CISER Data Archive Collection Policy: http://ciser.cornell.edu/pub/policies/CISER_Data_Collection_Policy.pdf [accessed 4/4/14]

CISER Data Archive Security Policy: http://ciser.cornell.edu/pub/policies/CISER_Data_Security_Policy.pdf
 [accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

CISER are in the process of reviewing its technical infrastructure with a view to mapping and translating existing tasks and functions to the OAIS reference model. This gap analysis exercise will identify functional archival processes that require attention, rationalize practice, inform policy decision making, and yield a more comprehensive understanding of archival standards that can be used to best deliver efficiency and trusted

repository status.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

14. The data consumer complies with access regulations set by the data repository.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

14. Data consumer complies with access regulations set by the data repository

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guidelines has been fully implemented for the needs of our repository.

Self-assessment statement:

CISER data consumers must agree to CISER Data Archive Use Policies (including the CISER Terms of Use Policy), CISER's system usage agreement, and observe CISER System Usage Policies. CISER computing account users are responsible for complying with all applicable federal, state and local laws, as well as Cornell University's policy in the use of CISER systems.

Users are responsible for selecting the appropriate level of technical security based upon the type of data that they use on CISER servers. Users also need to assert that they are in compliance with security safeguards required for the type of data for intended use.

The CISER Microsoft Windows environment employs user-based authentication. A user's access is authenticated by Kerberos using a unique username and password. A username serves to login to a compute server and/or download datasets from the CISER website where applicable.

Access to restricted data conforms to the requirements of Data Provider Agreements on a case by case basis.

Links to supporting documentation:

CISER System Usage Agreement - http://ciser.cornell.edu/computing/manual/useAgreement.shtm[accessed 4/4/14]

CISER System Usage Policies - http://ciser.cornell.edu/computing/manual/policies.shtm [accessed 4/4/14]

CISER Data Archive Use Policies - http://ciser.cornell.edu/info/policy.shtml [accessed 4/4/14]

CISER Terms of Use Policy: http://ciser.cornell.edu/pub/policies/CISER_Terms_of_Use.pdf [accessed 4/4/14]

CISER Data Archive Security Policy: http://ciser.cornell.edu/pub/policies/CISER_Data_Security_Policy.pdf [accessed 4/4/14]

Cornell University Policy Office (Policy Volume V. Information Technologies - http://www.dfa.cornell.edu/treasurer/policyoffice/policies/volumes/informationtech/index.cfm [accessed 4/4/14]

Cornell University IT Policy and Law - http://www.it.cornell.edu/policies/university/privacy/abuse/index.cfm [accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in higher education and scientific research for the exchange and proper use of knowledge and information.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

All data consumers of Cornell networks must abide by Cornell University policies which are in line with generally accepted higher education policies. In addition, any account holder of CISER's computing systems is provided a use agreement encouraging the client to use appropriate safeguards with accessing, storing and using any, and all,

data. All data consumers must agree to CISER Terms of Use Policy prior to downloading datasets.

Links to supporting documentation:

CISER Terms of Use: http://ciser.cornell.edu/pub/policies/CISER_Terms_of_Use.pdf [accessed 4/4/14]

Computing account agreement: http://ciser.cornell.edu/computing/manual/useAgreement.shtm[accessed 4/4/14]

CISER Systems Use Policies: https://ciser.cornell.edu/computing/manual/policies.shtm [accessed 4/4/14]

Cornell University Policy 4.12, Data Stewardship and Custodianship

http://www.dfa.cornell.edu/treasurer/policyoffice/policies/volumes/governance/data.cfm [accessed 4/4/14]

Cornell University Policy Library:

http://www.dfa.cornell.edu/treasurer/policyoffice/policies/volumes/index.cfm [accessed 4/4/14]

Cornell University Campus Code of Conduct: http://www.policy.cornell.edu/Campus_Code_of_Conduct.cfm [accessed 4/4/14]

Cornell University Policy Regarding Abuse of Computers and Network Systems
http://www.cit.cornell.edu/policy/responsible-use/abuse.html [accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

16. The data consumer respects the applicable licences of the data repository regarding the use of the data.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

16. Data consumer respects the applicable licenses of the data repository regarding the use of the data

Minimum Required Statement of Compliance:

4. Implemented: This guidelines has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guidelines has been fully implemented for the needs of our repository.

Self-assessment statement:

CISER data consumers must agree to CISER Data Archive Use Policies, the CISER's System Usage Agreement and observe CISER System Usage Policies. CISER computing account users are responsible for complying with all applicable federal, state and local laws, as well as Cornell University's policy in the use of CISER systems.

Users are responsible for selecting the appropriate level of technical security based upon the type of data that they use on CISER servers. Users need to assert that they are in compliance with security safeguards required for the type of data for intended use and also agree to adhere to licensing requirements as stipulated by the data provider.

Users agree to adhere to any and all licensing requirements as stipulated by the provider of datasets held in the CISER Data Archive.

Links to supporting documentation:

CISER Terms of Use: http://ciser.cornell.edu/pub/policies/CISER_Terms_of_Use.pdf [accessed 4/4/14]

CISER System Usage Agreement - http://ciser.cornell.edu/computing/manual/useAgreement.shtm [accessed 4/4/14]

CISER System Usage Policies - http://ciser.cornell.edu/computing/manual/policies.shtm [accessed 4/4/14]

CISER Data Archive Use Policies - http://ciser.cornell.edu/info/policy.shtml [accessed 4/4/14]

Cornell University Policy Office (Policy Volume V. Information Technologies)
http://www.dfa.cornell.edu/treasurer/policyoffice/policies/volumes/informationtech/index.cfm [accessed 4/4/14]

Cornell University IT Policy and Law - http://www.it.cornell.edu/policies/university/privacy/abuse/index.cfm [accessed 4/4/14]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: