



Implementation of the CoreTrustSeal

The CoreTrustSeal board hereby confirms that the Trusted Digital repository Cornell Institute for Social and Economic Research complies with the guidelines version 2017-2019 set by the CoreTrustSeal Board. The afore-mentioned repository has therefore acquired the CoreTrustSeal on April 13, 2018.

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

Yours sincerely,

The CoreTrustSeal Board

Assessment Information

Guidelines Version:	2017-2019 November 10, 2016
Guidelines Information Booklet:	CTS Requirements 2017-2019
All Guidelines Documentation:	Documentation
Repository:	Cornell Institute for Social and Economic Research
Seal Acquiry Date:	Apr. 13, 2018
For the latest version of the awarded CoreTrustSeal for this repository:	https://www.coretrustseal.org/why-certification/certified-repositories/
Previously Acquired Seals:	None
This repository is owned by:	Cornell Institute for Social and Economic Research

T 607-255-4801
F 607-255-9353
E ciser@cornell.edu
W <http://www.ciser.cornell.edu/Default.shtml>

Assessment

0. Context

Applicant Entry

Self-assessment statement:

Repository type: Subject-based, Institutional, Publication, Research project, National

Brief description of the Repository's Designated Community: 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.(1)

Level of Curation Performed:

D. Data-level curation – as in C above, but with additional editing of deposited data for accuracy.

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.

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.

Outsourcing of functions are limited to backup data storage (through E-Z Backup) and the hosting of our Dataverse instance using Amazon Web Services, as well as systems support from Cornell's Center for Advanced Computing(CAC).(2) We manage all aspects of support for the researcher throughout the data life cycle. This does not preclude CISER from utilizing tools that have been developed outside to enhance service delivery, such as StatTransfer for data conversion, Dataverse for preservation, and Sledgehammer for metadata creation. CISER is aligned with The Roper Center for Public Opinion Research(3), strategically sharing resources in IT and providing mutual support and expertise for data operations. The Roper Center for Public Opinion Research is a sustainable domain repository that has been reliably managing and providing access to public opinion data since 1947. In addition, CISER maintains formal relationships with the Inter-university Consortium for Political and Social Research (ICPSR), and the Data Documentation Initiative Alliance.

CISER is housed within Cornell's Office of the Vice Provost for Research (OVPR), which is dedicated to the support of research at Cornell. The OVPR, along with our member colleges, provides funding and strategic connections to sustain CISER's social science research support resources.(4) Colleges with primary and secondary connections to the social sciences provide financial support of our shared resources for their researchers. They are also strategic partners, providing input for planning for the resource needs of the future. Our member colleges include: College of Agriculture and Life Sciences, College of Arts and Sciences, College of Human Ecology, School of Industrial and Labor Relations, SC Johnson College of Business, College of Engineering and Computing and Information Science.(5) At present, the research archive contains nearly 2,000 studies with approximately 22,300 individual files totaling about 713 gigabytes plus 3 studies stored using compressed data totaling 210 GB. In addition, CISER holds a collection of more than 650 studies on CD and DVD.

Links to Supporting Documentation:

(all links visited 2/12/18)

1. About CISER: <http://ciser.cornell.edu/data/data-archive/>

2. Center for Advanced Computing: <https://www.cac.cornell.edu/>

3. Roper Center Board of Directors (see ex-officio board members at bottom, CISER director William Block): <https://ropercenter.cornell.edu/about-the-center/board-of-directors/>

4. Cornell University Research Division: <https://research.cornell.edu/research-division>

5. About CISER – Partners: <https://ciser.cornell.edu/about-us/partners/>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

1. Mission/Scope

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

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.(1)

CISER's mission(2) 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.

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 who work with data providers to ensure that data and accompanying documentation comply with CISER standards and policies.(3)(4)(5) CISER also provides access to a substantial web-based library of sample programs and instructional material on using the CISER research servers. (6)

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 professional organizations (See requirement #5). 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(7) and the New York Census Research Data Center(8) which provides academic researchers with access to selected Census confidential microdata in physically secure facilities.

CISER develops/monitors/updates its mission statement through an internal process involving all staff. CISER's senior leadership team finalizes and has ultimate responsibility for the mission statement, making sure it is aligned with the university mission, objectives and goals, and will lead to enhanced support of the needs of its member colleges and their researchers. Changes to the mission statement are shared with the Cornell Office of the Vice

Provost for Research and member colleges.

Links to supporting documentation:

(all links visited 2/12/18)

1. CISER history: <https://ciser.cornell.edu/about-us/history/>

2. CISER Data Archive Mission Statement:
https://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Mission_Statement.pdf

3. CISER Data Archive Preservation and Storage Policy:
https://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Preservation_and_Storage_Policy.pdf

4. CISER Data Archive Collection Policy:
https://ciser.cornell.edu/wp-content/uploads/2017/10/CISER_Data_Collection_Policy.pdf

5. CISER Terms of Use: https://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Terms_of_Use.pdf

6. CISER Available Software: <https://ciser.cornell.edu/computing/software/>

7. CRADC: <https://ciser.cornell.edu/data/secure-data-services/cradc/>

8. NYCRDC: <https://nyrdc.cornell.edu/>

Reviewer Entry

CoreTrustSeal Board

W www.coretrustseal.org

E info@coretrustseal.org

Accept or send back to applicant for modification:

Accept

Comments:

2. Licenses

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline 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 data consumers must observe CISER's Terms of Use(1) and System Usage Policies(2) and agree to CISER Data Archive Use Policies (3). 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.(4)(5)

As explained in requirement #9 below, the dissemination of data from the CISER Data Archive is built upon a "green-yellow-red" light system to signify whether data is publicly-available or restricted to Cornell researchers. Users are responsible for maintaining the appropriate level of technical security based upon the type of data that they use on CISER servers by asserting 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. Per CISER System Usage Policies, violations of this policy will result in the loss of access privileges.(2)

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.

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 Cornell's Office of Sponsored Programs (OSP)(6) and Institutional Review Board (IRB)(7) 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. (8)

Links to supporting documentation:

(all links visited 2/12/18)

1. CISER Terms of Use: https://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Terms_of_Use.pdf

2. CISER System Usage Policies - https://ciser.cornell.edu/wp-content/uploads/2017/11/CISER_Systems_Use_Policy.pdf

3. CISER Data Archive Use Policies - <https://ciser.cornell.edu/data/data-archive/>

4. Cornell University Policy Office (Policy Volume V. Information Technologies) <https://www.dfa.cornell.edu/policy-library> (click on “Information Technologies” box and apply)

5. Cornell University IT Policy and Law - <https://it.cornell.edu/policy/policy-50-abuse-computers-and-network-systems>

6. Office of Sponsored Programs: <https://www.osp.cornell.edu/>

7. Institutional Review Board: <https://www.irb.cornell.edu/>

8. Steps to Acquire and Use Restricted Data at Cornell: <https://ciser.cornell.edu/data/secure-data-services/cradc/>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

CoreTrustSeal Board

W www.coretrustseal.org

E info@coretrustseal.org

Comments:

3. Continuity of access

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

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.(1, Reason for Policy)

The CISER Data Preservation and Storage Policy documents the main theoretical and practical steps for providing long-term preservation of digital research data.(1) See requirement #10 for a description and discussion of CISERs data collection policy. (2)

CISER routinely monitors technical developments, as described in requirements# 10 and 14 of this document.(1, Management of Storage Infrastructure)

CISER operates within the framework of the Cornell University Continuity of Operations plan.(3) Working within this framework, CISER has created a plan specific to its needs and facility, which includes a recovery time objective of 24 hours for website access and a 3-day objective for archive access should there be a disaster or other failure of access.(4) Data continuity is also provided by Data-PASS, a voluntary partnership to archive, catalog, and preserve data used for social science research.(5) As mentioned elsewhere in this document, off-site backup of the archive is performed on a daily basis.

Links to supporting documentation:

(links visited 2/12/18)

1. CISER Data Archive Preservation and Storage Policy:
http://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Preservation_and_Storage_Policy.pdf

2. CISER Data Archive Collection Policy
http://ciser.cornell.edu/wp-content/uploads/2017/10/CISER_Data_Collection_Policy.pdf

3. Cornell University Continuity of Operations Plan:
https://emergency.cornell.edu/wp-content/uploads/EM_REC_COOPplan_v2_2017.pdf

4. CISER Continuity of Operations Plan (not public for security reasons)

5. Data-PASS: <http://www.data-pass.org/>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

4. Confidentiality/Ethics

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

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. (1, 2, 3) 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.(4, 5) All data consumers must agree to CISER Terms of Use Policy prior to downloading datasets.(6)

CISER makes efforts to confirm that data were collected in accordance with legal and ethical criteria in place at the time and place of its collection, especially review by Ethical or Institutional Review Boards (IRB). Where this information is unavailable, the professional judgement of the Data Librarian and the Director will be used to decide on the inclusion of such data, taking into account the relative risk (usually low) associated with the data.(7, "PURCHASE OF DATA DATASET APPRAISAL AND ACQUISITION") In the case of R-squared (CISER's Results Reproduction Service), CISER staff reproduce the findings to ensure the publication is accurately representative of the data.(8)

As mentioned previously, the majority of CISERs data files 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), where they are stored for a fixed amount of time, rather than in perpetuity, as described below.(9)

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.(10) The restricted access files are backed-up, unless specifically excluded in the Data Provider Agreement. The metadata is not searchable but rather is stored with the restricted data files and is only accessible by authorized and authenticated users. (11, particularly "Restricted Access Research Data Storage")

Restricted access files are kept on the CRADC secure file server, located in the university Data Center. The original restricted access data files supplied by the data providers are stored on physical media in a fireproof safe in the CISER building. (11, *ibid*) The CRADC secure file server utilizes self- encrypting disk (SED) and provides the storage media for both the original data files and the researchers working files. Backups of the data files are based on the Data Provider Agreements. In cases where backups are allowed, the files are encrypted in transport

and remain encrypted on the backup media. The data use agreements with data providers typically require that at the end of the project period the original media be returned or destroyed and that all copies of the data be destroyed. A data destruction certificate is provided to data providers at the end of the project. (12, "Destruction of Physical Media")

Data held by CRADC are categorized restricted-access and are only available via a legally signed contract with the Data Provider Agreements (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.(11, particularly "Data Center Specifications")

Procedures are in place to review disclosure risk in data, and to take the necessary steps to either anonymize files or to provide access in a secure way both in the CISER archive and in CRADC (where acceptable).(7, 11, 13)

CRADC Data Custodians take mandated training through CITI Program (14) in IRB Administration & Information Privacy & Security. Certificates are required to be renewed as per CITI requirements (5 years). Other trainings are included on a per-provider basis as outlined in data provider agreements - for example, The Bureau of Labor Statistics requires its own ongoing training required to be renewed every 3 years.(15) A CISER Research Associate has attended ICPSR training in Assessing and Mitigating Disclosure Risk. (16) For users, guidance in the responsible use of disclosive data is provided under the Data Provider guidelines for CRADC, but is not necessary for the CISER Research archive.

CISER reserves the right to disable a computing account immediately upon identification of possible misuse of any CISER services. Account termination will occur if misuse is confirmed through proper authorities, and no reinstatement will be allowed.(5, 6)

Links to supporting documentation:

(All links visited 2/12/18)

- 1..Cornell University Policy Library: <https://www.dfa.cornell.edu/policy>
2. Cornell University Campus Code of Conduct: <https://www.dfa.cornell.edu/sites/default/files/policy/CCC.pdf>
3. Cornell University Policy Regarding Abuse of Computers and Network Systems
<https://it.cornell.edu/policy/policy-50-abuse-computers-and-network-systems>
4. Computing account agreement: <https://ciser.cornell.edu/computing/>
5. CISER Systems Use Policy:
https://ciser.cornell.edu/wp-content/uploads/2017/11/CISER_Systems_Use_Policy.pdf

6. CISER Terms of Use: http://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Terms_of_Use.pdf
7. CISER Data Archive Collection Policy:
http://ciser.cornell.edu/wp-content/uploads/2017/10/CISER_Data_Collection_Policy.pdf
8. CISER: Data Curation and Reproduction of Results Service:
<https://ciser.cornell.edu/research/results-reproduction-r-squared-service/>

9. What is CRADC?: <https://ciser.cornell.edu/data/secure-data-services/cradc/>

10. CRADC Access Control Policy:
http://ciser.cornell.edu/wp-content/uploads/2017/01/CRADC_AccessControlPolicy_Provisioning_and_Deprovisioning.pdf

11. CRADC Data Security Policy:
http://ciser.cornell.edu/wp-content/uploads/2017/01/CRADC_Data_Security_Policy.pdf

12. CRADC Data Destruction and Return of Restricted Data Policy:
http://ciser.cornell.edu/wp-content/uploads/2017/01/CRADC_Destruction_and_Return_of_Restricted_Data.pdf

13. Cornell University Policy 4.12, Data Stewardship and Custodianship:
https://www.dfa.cornell.edu/sites/default/files/policy/vol4_12.pdf

14. CITI Program: <https://about.citiprogram.org/en/homepage/>

15. BLS Restricted Data Access: <https://www.bls.gov/rda/faqs.htm#q9>

16. Assessing and Mitigating Disclosure Risk: Essentials for Social Science:
<https://www.icpsr.umich.edu/icpsrweb/sumprog/courses/0115>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

5. Organizational infrastructure

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

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.(1)

CISER is funded through Cornell University. The university operates on an annual budget model, although it is planned out for multiple years. Based on the ongoing model of funding, Cornell has sufficient funding to maintain CISER's current level of staffing and IT resources and for attending meetings and training for the next 3 – 5-years.(2) CISER currently has 13 staff with an FTE total of 1130%.

Cornell provides training and professional development opportunities which are shared through regular communications with all employees, as well as other opportunities that are shared with supervisors to offer their employees in specific areas. The leadership at CISER recognizes the importance of and actively pursues opportunities for training and professional development in general and for subject matter expertise. Leadership keeps track of external opportunities and encourages attendance and sends staff. Leadership is an active participant in many organizations that offer these opportunities, such as IASSIST and NADDI.(3)

CISER staff have the breadth of knowledge necessary for a Social Science and Economics archive. CISER staff hold doctorates in History and Development Sociology(multiple) and master's degrees in Management, Sociology, Development Sociology, History(multiple), Library and Information Science, and Demography. CISER and CISER staff belong to several organizations including ICPSR, COPAFS, DDI, APDU, and Educause (as Organization members), as well as PPA, IASSIST, PAA, AAPOR, the American Sociological Association, SSHA, IUSSP, the American Statistical Association, and ISSA individually. The Director is past President of IASSIST (4, "Ex-Officio Officers") , and currently Executive Director of the Social Science History Association(5). The Senior Research Associate is Immediate Past President of the Association of Public Data Users(6), served on the Steering Committee of the American Community Survey Research Group, and is a member of the Standing Committee on Reengineering Census Operations(7). The Research Associate is co-chair of the International Fellows Organization of IASSIST(8) and is Cornell Organizational Representative to ICPSR.

Relevant links:

CoreTrustSeal Board

W www.coretrustseal.org

E info@coretrustseal.org

(all links visited 2/12/18)

1. Office of the Vice Provost for Research – Affiliated Research Centers:
<https://ovpr.research.cornell.edu/centers.html>

2. Cornell University Budget Office: <http://dbp.cornell.edu/home/offices/university-budget-office/>

3. About CISER: <http://ciser.cornell.edu/about-us/partners/>

4. IASSIST Officials: <http://www.iassistdata.org/about/officials.html>

5. SSHA Current Officers: <https://ssha.org/officers/>

6. APDU Board of Directors: <http://apdu.org/about-apdu/board-of-directors/>

7. PAD People: <https://pad.human.cornell.edu/people.cfm>

8. IASSIST Fellows Committee: <http://www.iassistdata.org/about/outreach.html>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

6. Expert guidance

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

There are multiple mechanisms CISER uses to get oversight, input, and guidance to improve performance. These mechanisms include scheduled face-to-face meetings with the academic and administrative leadership of each member college; meetings with individual faculty members identified in multiple ways (self-identified, suggested by colleges, selected from user activity); active participation in multiple groups that are involved in the service areas CISER supports (from Cornell-based units such as the Research Data Management Service Group (described below) to topic-specific entities, as well as national and international organizations).

CISER is a member of the Cornell University Research Data Management Service Group (RDMSG), a collaborative, campus-wide organization that links Cornell University faculty, staff and students with data management services to meet their research needs. The RDMSG's broad range of science, policy, data, and information technology experts provide timely and professional assistance for the creation and implementation of data management plans, and help researchers find specialized management services they require at any stage of the research process, including initial exploration, data gathering, analysis and description, long term preservation and access.(1) The CISER Director is a member of the Coordinator and Management Council of the RDMSG.(2) The CISER Research Associate is a consultant for RDMSG.(3)

For consultations, RDMSG consultants are available for consultation weekdays from 2-3pm and at other times upon request, via phone, by email or in person.(1, "Consultations")

For users, CISER provides a Help Desk service Monday through Friday 8:30am-10:30am and 12:30pm-6:00pm. This service is available online, by phone, email, or walk-in.(4) It is fully available for all members of the Cornell community, as well as others who need assistance in finding data within the CISER catalog.

Links to supporting documentation (all links visited 2/12/18):

1. Cornell University Research Data Management Group: <https://data.research.cornell.edu/>
2. RDMSG Coordinator and Management Council: <https://data.research.cornell.edu/content/coordinator-and-management-council>
3. RDMSG Consultants: <https://data.research.cornell.edu/content/consultants>
4. CISER Help Desk Services: <http://www.ciser.cornell.edu/computing/HelpDesk.php>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

7. Data integrity and authenticity

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

DATA INTEGRITY:

A checksum is generated by using a 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. (1, "Data Integrity") MD5 checksum validations are run to report all files which have been added, deleted, or modified since the previous validation.(1, "Data Integrity," "Data Normalization")

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

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

o List which researchers have access to restricted files. (2, 3)

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

The determination of completeness of metadata is made at the discretion of the Data Librarian. (1, "Responsibilities" and 4, "Responsibilities") Lists of studies which are missing digitized codebooks and other data have been made, and efforts are currently underway to digitize physical codebooks for those studies which are missing them in the archive. Changes to the data themselves, or major changes to metadata, are issued as a new version of the dataset.(1, "Data Normalization"), (5)

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.(5)

CISER is committed to following Data Documentation Initiative standards, as well as the Open Archival Information Systems Reference Model, Data Seal of Approval, and Trusted Repositories Audit and Certification.(1)

AUTHENTICITY MANAGEMENT:

CISER staff correct any errors in the data (in collaboration with the data provider) found on our end (which can be found by CISER staff or reported by users (see requirement #11.) (5) Users can report data errors by filling out a form on the website reached by clicking on a “Report an Error” button at the bottom of each study record (6) or by contacting the help desk.

The archive database includes fields for the Primary Investigator of a study, the producer, distributor, and the source of the data, that is, from where CISER received it. When CISER replaces a file with a newer version, CISER staff enters this information in the file’s database record.(5)

Documentation and codebooks are provided along with the data when available. Related datasets are given the same or similar codebook numbers (for example, Current Population Survey: School Enrollment are cataloged as CPH-010(1968) through CPH-010(1985)). An in-house subject scheme is used to allow users to browse through similar datasets.(7)

As described in requirement #9, 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.(1)

Where possible CISER will clearly label and make available earlier versions of data and documentation through the data catalog. Version record numbers are captured in metadata held in CISER relational databases. CISER retains the right to withdraw an older version of a data study where significant change may be

misrepresentative.(5)

The vast majority of CISER's holdings are purchased or freely available datasets from external providers. With the implementation of the R-squared service (discussed in requirement #4), it is expected that the proportion of Cornell-created datasets should rise – this is currently a percentage of the archive in the single digits. In most cases, these depositors are personally known to the Institute and files are provided and entered into the archive by CISER staff. In the future, plans are to use specific log-in usernames and passwords to authenticate deposit based on a data deposit model developed jointly with the Roper Center and in early implementation stage at Roper. (8)

Links to supporting documentation:

(all links visited 2/12/18)

1. CISER Data Archive Preservation and Storage Policy:

https://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Preservation_and_Storage_Policy.pdf

2. CISER Data Archive Security Policy:

http://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Security_Policy.pdf

3. CRADC Data Security Policy:

http://ciser.cornell.edu/wp-content/uploads/2017/01/CRADC_Data_Security_Policy.pdf

4. CISER Data Archive Collection Policy:

http://ciser.cornell.edu/wp-content/uploads/2017/10/CISER_Data_Collection_Policy.pdf

5. CISER Data Archive Versioning Policy:

https://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Versioning_Policy.pdf

6. Report an error example: <https://cisermgmt.cornell.edu/go/PHPs/dataReportErrors.php?IDTITLE=2798>

7. Browse Data Archive Holdings by Subject: <https://cisermgmt.cornell.edu/go/PHPs/browse.php>

8. Roper Center Data Deposit: <https://ropercenter.cornell.edu/polls/deposit-data/>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

8. Appraisal

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

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 Collection Policy (1) drives data appraisal and acquisition for the CISER research archive. 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. 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, and expected future use by a broad constituency of social science users.

Using the same criteria, data are also acquired for students of those faculty who are engaged in substantive social science or economic research. Proactive collection development is undertaken in anticipation of demand. Criteria include the quality of the data and the reliability of the distributor, and expected future use by a broad constituency of social science users.

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 two weeks on the local server and six months on tape by the EZ-Backup service.) 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 validate data integrity, scripts are run on a scheduled basis, as described in requirement #9. (2, "Data Integrity")

Where possible, data are accompanied by comprehensive machine-readable documentation: codebooks, file layout maps, technical notes, questionnaires, reports, and errata in open and accessible formats. In cases where documentation is incomplete, the archive staff work with data producers to gather more, to ensure that data files are usable and understandable. The Data Archive reserves the right to reject datasets deemed inadequately documented. (1, "Documentation," "Data Quality") Metadata creation continues across the data lifecycle. If necessary, additional user information is 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.(3)

As mentioned earlier, metadata creation continues throughout the data life cycle, and additional user information will be provided when appropriate. (1, "Documentation" and "Data Quality", 3)

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. (1, "File Formats") 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.(1, 2, "Data Integrity") 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. (1, "File Formats")

Links to supporting documentation:

(all links visited 2/12/18)

1. CISER Data Archive Collection Policy:

http://ciser.cornell.edu/wp-content/uploads/2017/10/CISER_Data_Collection_Policy.pdf

2. CISER Data Archive Preservation and Storage Policy:

http://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Preservation_and_Storage_Policy.pdf

3. CISER Data Archive Versioning Policy:

http://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Versioning_Policy.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

9. Documented storage procedures

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The processes and procedures for managing archival storage are documented in the CISER Data Archive Preservation and Storage Policy which is linked below. (1) They are managed through a coordinated effort between the CISER data librarian, the CISER IT Director, and other CISER staff. Policy changes are approved by the CISER director.

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 (this is usually data provided by outside private providers such as ICPSR, Roper, and other providers from whom the data is purchased with an institutional license), while those classified with a “red light” are restricted and require permission prior to use as stipulated by respective data providers.(2) Users also need to assert that they are in compliance with security safeguards required for the type of data for intended use.(3)

The CISER Data Archive is stored on network attached storage (NAS) in both compressed and uncompressed format in Cornell University’s Data Center with 187TB capacity. 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. (4, “Data Center Specifications”, 5)

An entirely separate domain and servers are built specifically for restricted- use datasets, as documented in an Information System Security Plan (confidential document, can provide informational copy) per NIST 800-18 guidelines.(6, 7) 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.(8) 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. (9) CISER works with data providers to implement security plans meeting provider requirements. Part of this process is to review for disclosure risks.

Restricted access files are kept on the CRADC secure file server, located in the university Data Center. The original restricted access data files supplied by the data providers are stored on physical media in a fireproof safe

in the CISER building. (4, particularly "Restricted Access Research Data Storage") The CRADC secure file server utilizes self-encrypting disk (SED) and provides the storage media for both the original data files and the researchers working files. Backups of the data files are based on the Data Provider Agreements. In cases where backups are allowed, the files are encrypted in transport and remain encrypted on the backup media. The data use agreements with data providers typically require that at the end of the project period the original media be returned or destroyed and that all copies of the data be destroyed. A data destruction certificate is provided to data providers at the end of the project. (10, "Destruction of Physical Media")

CISER research archive 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.(11) See above for CRADC backup strategy.

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. (1, "Management of Storage Infrastructure") All CRADC backups are deleted at project close-out.(10)

CISER mitigates risk by using a fail-safe design on both a short-term and long-term basis. On a daily basis, EZ-Backup protects the archive, as outlined above. In addition, CISER is a member of DATA-Pass, a long-term partnership technique whereby another partner member agrees to take the data of another partner which loses its ability to maintain an archive (through loss of funding or similar reasons).(12)

To ensure that the digital content remains identical and accessible across archival copies, 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. (1, "Data Integrity")

Hardware Lifecycle Management principles are in place to maintain up-to- date systems and follow regular maintenance procedures.

Links to supporting documentation:

(links visited 2/12/18)

1. CISER Data Archive Preservation and Storage

Policy: http://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Preservation_and_Storage_Policy.pdf

2. CISER Data Archive Policies: <http://ciser.cornell.edu/data/data-archive/>

3. CISER system usage policies: <http://ciser.cornell.edu/computing/>

4. CRADC Data Security

Policy: http://ciser.cornell.edu/wp-content/uploads/2017/01/CRADC_Data_Security_Policy.pdf

5.

Minfo32.exe

OS: Microsoft Windows 2016 DataCenter

Hardware: Dell M830 and M630 Blade servers; CPU and GPU configurations

Additional hardware: R810 and R820 servers

Virtualization: Microsoft Hyper-V

Disk storage: 187 Tb

6. CRADC Access Control

Policy: http://ciser.cornell.edu/wp-content/uploads/2017/01/CRADC_AccessControlPolicy_Provisioning_and_Deprovisioning.pdf

7. Security and Privacy Controls for Federal Information Systems and Organizations:

<https://csrc.nist.gov/csrc/media/publications/sp/800-53/rev-5/draft/documents/sp800-53r5-draft.pdf>

8. Cornell DropBox (requires login): <https://dropbox.cornell.edu/login/>

9. Sharing, Transmission and Distribution of Restricted

Data: http://ciser.cornell.edu/wp-content/uploads/2017/01/CRADC_Sharing_Transmission_Distribution_Policy_1.pdf

10. CRADC Data Destruction and Return of Restricted Data

Policy:

http://ciser.cornell.edu/wp-content/uploads/2017/01/CRADC_Destruction_and_Return_of_Restricted_Data.pdf

11. CISER Data Archive Security

Policy: http://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Security_Policy.pdf

12. Data-PASS – About: <http://www.data-pass.org/about.jsp>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

10. Preservation plan

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

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. (1)

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. (2)

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.(1, “Management of Storage Infrastructure”)

CISER is committed to preserving and making available for use primary and secondary data, documentation and metadata, in discipline recognized digital formats that remain suitable for research in perpetuity. Long-term data preservation is integrated into archival operations and planning within CISER.(1)

Contracts differ in language and requirements depending on the distributor and method of acquisition. CISER uses a red-yellow-green light system to track access requirements.(3) Datasets with restricted information are housed in the CRADC archive and terms & requirements are tracked with Agiloft software.

Data transfer methods vary depending on the source of the acquisition. In the case of public-use data purchases requested by or created by Cornell researchers, data can be transferred on physical drives, optical discs, or through email or Dropbox services. In the case of restricted use data sets, data is transferred to the CRADC server through secure Dropbox services or on physical drives through FedEx delivery, or the courier of the provider’s choice.(4)

CISER retains rights to copy, transform, and store the data in accordance with data provider agreements, unless the data are embargoed, in which case the files are stored in the “red-light” category (no access without authorization) until access is allowed.(3)

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.(1)

Measures to ensure these actions are taken currently reside with the Data Librarian. however, our data deposit application is being revamped. Through joint efforts with The Roper Center and CISER, software engineers have designed and developed the data deposit application entirely internally. The application is in the early stages of implementation at Roper. CISER plans to adopt the process with slight modification shortly. (5)

Supporting Documentation:

(Links visited 2/12/18)

1. CISER Data Archive Preservation and Storage

Policy: http://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Preservation_and_Storage_Policy.pdf

2. CISER Data Archive Collection

Policy: http://ciser.cornell.edu/wp-content/uploads/2017/10/CISER_Data_Collection_Policy.pdf

3. CISER Data Archive Policies: <http://ciser.cornell.edu/data/data-archive/>

4. Sharing, Transmission and Distribution of Restricted Data:

http://ciser.cornell.edu/wp-content/uploads/2017/01/CRADC_Sharing_Transmission_Distribution_Policy_1.pdf

5. Roper Center Data Deposit: <https://ropercenter.cornell.edu/polls/deposit-data/>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

11. Data quality

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

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.(1)

CISER adheres to DDI Lifecycle standards for metadata.(2) The SQL database is mappable to DDI at the study level. CISER is in the process of creating a Dataverse instance which has five required fields and maps to DDI, as well as generating DDI XML exports.(3)(4)

CISER accepts feedback on all user issues including data and metadata quality through email or the help desk using a Remedy ticketing system, and informally at meetings and conferences. (5)(6) When known to exist, citations and links to related works such as journal articles are provided to aid in data sharing and discovery of prior publications and findings using the dataset. (See 7 for an example) Acknowledgement of use of CISER support and resources is requested with a provided acknowledgement statement.(8)

Links to supporting documentation:

(all links checked 2/12/18)

1. CISER Data Archive Collection Policy:

http://ciser.cornell.edu/wp-content/uploads/2017/10/CISER_Data_Collection_Policy.pdf

2. CISER Data Archive Preservation and Storage

Policy: http://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Preservation_and_Storage_Policy.pdf

3. The Dataverse Network: <https://www.ddialliance.org/project/the-dataverse-network>

4. Example of a DDI XML export from the CISER Dataverse can be created by clicking on the metadata tab and “Export Metadata”: <https://dataverse.cornell.edu/dataset.xhtml?persistentId=doi:10.6077/J5/BKK8DB>

5. CISER Help Desk Services: <http://ciser.cornell.edu/consulting/ciser-helpdesk/>

6. Remedy: <https://it.cornell.edu/remedy>

7. Attractive Names Sustain Increased Vegetable Intake in Schools:
https://cisermgmt.cornell.edu/go/ASPs/search_athena.asp?IDTITLE=2797

8. Acknowledgement: <https://ciser.cornell.edu/acknowledgement/>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

12. Workflows

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

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) (1) (2, particularly “File Formats” and “Documentation”)

CISER staff who manage data have a set of internal guidelines that they adhere to and they document ingest processes and data transformations. (3) Other processes such as long-term preservation (e.g. normalization, version control, sustainability) are detailed in the CISER Data Preservation and Storage Policy and CISER Data Archive Versioning Policy. (4)

User direction is provided in the CISER System Usage Policies (5) which must be agreed to before an account is provided. These directions include but are not limited to direction that CISER accounts are for research/academic use only, not used for personal purposes, and that passwords may not be shared with any other person. This information is shown in a statement whenever the user logs into the CISER research server.

Additional permission settings need to be created on the server for studies which require authorization from the data provider. (See 6, "Use of Archive Data," particularly the "red-light" description) After processing, a checksum is run on the web archive to ensure all files are uploaded and appear in the file listing for the study. (4)

The CISER Data Collection Policy (2) details criteria and information regarding the selection of data for archiving. 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.

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. (2, "Confidentiality")

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. (2) As described below, CISER will accept data regardless of physical format as long as they are convertible to supported and accessible file formats suited for long-term preservation for use by the entire Cornell community. In these cases, CISER staff will normalize data into accompanying raw ASCII or Unicode.(2)

In some cases, now rare, text documentation or program files have no carriage returns, in these cases, a Unix2Dos utility is run which creates a carriage return at the end of each record to display properly. If files are restricted, permissions are added at the folder-level (when all files in the folder receive the same permissions) or at the file-level (when different permissions apply to files within the same folder).(3)

Data acquisition policies have been covered previously and appear in the CISER Data Archive Collection policy. The CISER Director and CISER Data Librarian are tasked with interpreting these policies, provide clarification and education, and implement operational and business processes to facilitate compliance.(2)

CISER routinely monitors appropriate changes and improvements in technology and our users' needs. CISER staff discuss, agree to, and implement updates to policies and workflows when such changes are deemed necessary or desirable. (4, "Management of Storage Infrastructure")

Supporting documentation:

(all links visited 2/12/18)

1. CISER Mission Statement: https://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Mission_Statement.pdf

2. CISER Data Archive Collection

Policy: https://ciser.cornell.edu/wp-content/uploads/2017/10/CISER_Data_Collection_Policy.pdf

3. Processing new archive data on the CISER Research Computing System (internal document, can be provided).

4. CISER Data Archive Preservation and Storage Policy:

https://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Preservation_and_Storage_Policy.pdf

5. CISER System usage policies:

https://ciser.cornell.edu/wp-content/uploads/2017/11/CISER_Systems_Use_Policy.pdf

6. CISER Data Archive Policies: <https://ciser.cornell.edu/data/data-archive/>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

CoreTrustSeal Board

W www.coretrustseal.org

E info@coretrustseal.org

13. Data discovery and identification

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

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 online catalog is listed on the Registry of Research Data Repositories located at re3data.org(1) and 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. 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. (2 (particularly "Data Normalization")) (3) (4) 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), and Data Documentation Initiative (DDI)). (2, "Reason for Policy")

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, Mathematica, Matlab, R, SAS, SPSS, Stata) For complete list see Available Software.(5) The documentation provided with most studies includes a standard format study- level citation and bibliographic information is provided with each dataset. (See 6 for an example.)

The CISER Data Archive does not currently have the technology in place to facilitate machine harvesting of the metadata. However, once the archive is fully migrated to Dataverse there will be the capability to allow for machine harvesting of the metadata.(7) While CISER has no plans to create an API, Dataverse is fully machine harvestable.(See requirement #15)

All data studies maintained by the CISER Data Archive are assigned a locally-generated unique identifier. (see https://cisermgmt.cornell.edu/go/ASPs/search_athena.asp? IDTITLE=2775 for example, where IDTITLE=2775 serves as the unique identifier) CISER is currently in the process of creating a Dataverse repository which will assign unique DOIs to each dataset. (7) CISER is a member of EZID.(8, "Current Users")

Supporting documentation:

CoreTrustSeal Board

W www.coretrustseal.org

E info@coretrustseal.org

(Links visited 2/12/18)

1. Re3data.org record for CISER: <http://www.re3data.org/repository/r3d100011056>

2. CISER Data Archive Preservation and Storage Policy: http://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Preservation_and_Storage_Policy.pdf

3. CISER Data Archive: Online catalog: <https://cisermgmt.cornell.edu/go/ASPs/search.asp>

4. Browse Data Archive Holdings by Subject: <https://cisermgmt.cornell.edu/go/PHPs/browse.php>

5. Available software: <http://ciser.cornell.edu/computing/software/>

6. Census of Population and Housing, 1980: Public Law 94-171 Population Counts, New York: https://cisermgmt.cornell.edu/go/ASPs/search_athena.asp?IDTITLE=60

7. CISER Dataverse: <https://dataverse.cornell.edu/dataverse/CISER>

8. EZID Partners & Clients: <https://ezid.cdlib.org/learn/>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

14. Data reuse

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

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 (DDI, OAIS, etc.). 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. (1)

Data are provided primarily in ASCII format with access to statistical software packages (SAS, SPSS, Stata, R, Matlab, Stat Transfer and others), as well as programming utilities. Some datasets already include SAS, SPSS, Stata, etc. files.(1) (2 (“Reason for Policy,” Data Normalization”)) During the ingest process, non-supported file formats are converted to these formats. 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. (1) (2) (3 (“Reason for Policy,” Data Normalization”)) When available documentation is stored on-site and in digital formats. (1) Efforts are underway to convert physical documentation to digital format for those studies which do not yet include both. The software packages mentioned earlier (SAS, SPSS, Stata, R, Matlab, Stat Transfer and others) are provided to make data more understandable and easy to work with. (4)

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. (2, “Management of Storage Infrastructure”)

Links to supporting documentation:

(all links checked 2/12/18)

1. CISER Data Archive Collection

Policy: https://ciser.cornell.edu/wp-content/uploads/2017/10/CISER_Data_Collection_Policy.pdf

2. CISER Data Archive Preservation and Storage

Policy: https://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Preservation_and_Storage_Policy.pdf

3. CISER Data Archive Versioning

Policy: https://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Versioning_Policy.pdf

4. Available Software: <https://ciser.cornell.edu/computing/software/>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

15. Technical infrastructure

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

CISER reference standards are primarily W3C for the majority of its services but does offer ArcGIS on GPU Processors.(1) Windows OS is updated monthly and hardware is updated on a regular basis of 3-4 years. Applications are updated on an annual basis. W3C is delivered via an IIS Server. There is no significant deviation from the standard. (2, see “Hardware Systems”)

CISER selected to implement a Dataverse instance (3) based on the standards it offered, along with its breadth of capabilities. Dataverse is an open-source software to share, cite, and find data. Dataverse provides for the use of DDI, Dublin Core, and JSON metadata standards. In addition, it is capable of citing datasets in EndNote XML, RIS, and BibTeX standards. If settings are defined and activated, Dataverse will also allow metadata harvesting through the utilization of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) by exposing the structured metadata.

Furthermore, within Dataverse, each record is associated with a Digital Object Identifier (DOI), a given persistent identifier. CISER’s DOIs are currently being provided through EzID.(4) EzID is a service from University of California at Berkley, which makes it simple for digital object producers to obtain and manage DOIs for their digital content. EzID was developed and is supported by the UC Curation Center (UC3). Over the next few months, EzID is phasing out and CISER, along with the Cornell University Library, will be purchasing a membership with DataCite.(5) DataCite, similar to EzID, provides DOIs as persistent identifiers to help the research community locate, identify, and cite research data.

To align with our mission, CISER resolves to maintain a high-quality state- of-the-art infrastructure (both servers and storage) and implements technological improvements on a 3-5 year hardware life cycle. Prior to selecting the next-generation infrastructure, options are thoroughly researched and identified to ensure CISER will benefit from technological advancements, improved performance and meeting growth requirements. Technology selection and acquisition strategy are created from the results of the research, with particular attention given to streamlining and enhancing our data preservation practices. Final selection of the infrastructure is designed to accommodate scalability, reliability, and sustainability, in accordance with quality control specifications and security regulations. Integration and implementation of the infrastructure is provided onsite or remotely by the vendor certified experts to validate operability in our environment. Post implementation, the infrastructure is covered by varying levels of support services, including proactive maintenance levels based on criticality of the resources to ensure operation is optimal at all times. As CISER continues to buildout a Dataverse instance, the metadata and files are stored on-site at Cornell University and in the cloud on an AWS EC2 server. The EC2 server hardware is maintained by Amazon Web Services and managed by CISER staff.(6) CISERs network is running on a 10-gigabyte backbone which is

based on Cornell Information Technology's network service (which runs on a 100- gigabyte backbone). (1)(7)

Software inventory is maintained by Cornell's Endpoint Management via Microsoft System Center Configuration Manager (SCCM).(8) Microsoft OneNote is used to manage documentation for IT internals. The software inventory is also displayed on CISER's website. (1) 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. Among the community-supported software available are R, LaTeX, Notepad++, and Python. (1) 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, R, Matlab, Stat Transfer); and a separate, secure computing environment to support use of confidential datasets.

Links to supporting documentation:

(Links visited 2/12/18)

1. CISER Available Software: <https://ciser.cornell.edu/computing/software/>
2. CISER Computing Resources: <https://ciser.cornell.edu/computing/>
3. CISER Dataverse: <https://dataverse.cornell.edu/dataverse.xhtml?alias=CISER>
4. EZID: <https://ezid.cdlib.org/>
5. Datacite: <https://www.datacite.org/>
6. CISER Data Preservation and Storage Policy:
https://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Preservation_and_Storage_Policy.pdf

7. Cornell Wired Network: Faculty and Staff - <https://it.cornell.edu/wired>

8. Endpoint Management Tools: <https://it.cornell.edu/endpoint-mgmt>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

16. Security

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

Archive data is backed up daily by Cornell Information Technology EZ- Backup service (using Tivoli Storage manager (TSM)). Metadata is backed up through SQL Server jobs nightly and is held for two-weeks before deleting. In addition, the metadata is backed up daily via the EZ-Backup service. A MD5 checker logs any changes to data. (1, "Management of Storage Infrastructure")

EZ-Backup accommodates for three copies of changed files in the backup database at all times. Deleted files remain available for 180-days, with plans to extend this time period to 375 days. The EZ-Backup service provides for an off-site replicate. 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. (1, particularly "Management of Storage Infrastructure")

The CISER servers are located in the CIT server farm, an environmentally controlled secure Data Center located at 757 Rhodes Hall, at Cornell University, Ithaca, NY. A proximity card reader secures the Data Center, and access is granted only to Cornell staff with the required credentials according to Cornell University Policy 8.4 – Management of Keys and Other Access Control Systems. Entrance and exits to the Data Center are automatically logged and monitored by Cornell Information Technology Staff.(2)(3) All CISER file servers have System Center Endpoint Protection - Windows Defender software installed, and data files are scanned for viruses prior to being added to the environment.(2, "Policy Guidelines")

CISER staff are mainly located at the CISER building at 391 Pine Tree Road, Ithaca, NY 14850. CISER offices use an authorized proximity card reader issued only to Cornell staff with the required credentials according to Cornell University Policy 8.4, referenced above.(3) Entrance to the CISER staff offices are automatically logged and monitored by CISER staff responsible for operation of the B.A.S.I.S. door security system.

Access to the Data Archive digital collection is preserved through Microsoft Windows NTFS permissions. Any original media/electronic data that is retained, will be stored in compliance with the CISER Data Archive Preservation and Storage Policy.(1) Reporting security incidents is mandated by Cornell University Policy 5.4.2, Reporting Electronic Security Incidents. (4)

CISER uses Solarwinds Log & Event Manager to provide compliance reporting, real-time event correlation and remediation, and file integrity monitoring.(5)

Key security personnel include IT Director & Security Liaison Janet Heslop and Software Engineer Brandon Kowalski. In addition, CISER contracts with the Center for Advanced Computing for System Administration support lead by Resa Reynolds, Assistant Director of Systems. (6)(7)

Links to supporting documentation:

(Links visited 2/12/18)

1. CISER Data Archive Preservation and Storage

Policy: https://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Preservation_and_Storage_Policy.pdf

2. CISER Data Archive Security

Policy: https://ciser.cornell.edu/wp-content/uploads/2017/01/CISER_Data_Security_Policy.pdf

3. Cornell University Policy 8.4 – Management of Keys and Other Access Control Units (see pp. 8-18 primarily) -

https://www.dfa.cornell.edu/sites/default/files/policy/vol8_4.pdf

4. Cornell University Policy 5.4.2 – Reporting Electronic Security Incidents:

https://www.dfa.cornell.edu/sites/default/files/vol5_4_2.pdf

5. Solarwinds Log & Event Manager: <http://www.solarwinds.com/log-event-manager>

6. CISER Staff Directory: <https://ciser.cornell.edu/contact-ciser/staff-directory/>

7. Center for Advanced Computing Staff Directory: <https://www.cac.cornell.edu/contact/directory.aspx>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

17. Comments/feedback

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

0. N/A: Not Applicable.

Self-assessment statement:

All links, including in requirements marked “accept,” have been updated to reflect CISER's new website. Some changes have been made to requirements marked “accept” to remove guidance text, to reflect information duplicated in other requirements, or for clarity. In addition, some issues were raised over a mention of “externally held datasets” in the CISER archive. This was in error and has been removed, CISER does not currently host any externally held datasets in the CISER archive. The CISER continuity plan which was described as being created has been finalized. As mentioned in the links, it is not public and can be provided to the reviewer if necessary, please provide an email.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: