



Assessment Information

[CoreTrustSeal Requirements 2020–2022](#)

Repository: Cornell Institute for Social & Economic Research
Website: <https://ciser.cornell.edu/>
Certification Date: 01 June 2021

This repository is owned by: **Cornell University**



Cornell Institute for Social & Economic Research

Notes Before Completing the Application

We have read and understood the notes concerning our application submission.

True

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Reviewer 3

Comments:

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background & General Guidance

Glossary of Terms

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type. Select all relevant types from:

Domain or subject-based repository, Institutional repository, Research project repository

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Accept

Brief Description of Repository

The CISER Data and Reproduction Archive has an extensive data collection with emphasis on the social sciences, including labor, political, and social behavior, demography, economics, and health.(1) The collection includes federal and state census files, administrative records, public opinion surveys, economic and social data from national and international organizations. In addition, the CISER Data and Reproduction Archive has studies compiled by individual researchers primarily from the Cornell community and collaborators and code/data packages created by our Results Reproduction Service (R-Squared). The goals of the CISER Data and Reproduction Archive are to make data available to researchers, while at the same time making data more Findable, Accessible, Interoperable, and Reusable (FAIR).(2)

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Accept

Brief Description of the Repository's Designated Community.

CISER's collection of data files emphasize studies that match the interests of Cornell social science researchers, staff, students, and faculty.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:
Accept

Level of Curation Performed. Select all relevant types from:

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

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:
Accept

Comments

CISER works with a range of national and international organizations as well as individual researchers to receive their data and code files. The majority of our studies are public use files that are placed in the CISER Data and Reproduction Archive for widespread access with a small number of studies that are restricted access based on licensing agreements. CISER employs a modified version of the Data Curation Network's CURATED workflow for ingest processing to ensure the quality and integrity of datasets.(3)

CISER works with data providers to resolve any missing information, inconsistencies, or 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 or documentation. Hard copies of the documentation are converted into PDF/A formats for archival and downloading purposes.

After the initial processing of the package, curation continues across the data life cycle (i.e., from data conceptualization to collection, processing, distribution, discovery, analysis, repurposing, and archiving). Additional user information will be provided as necessary, 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. In addition, annotations are indicated in the appropriate fields in the archive database.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:
Accept

Insource/Outsource Partners. If applicable, please list them.

In order to provide service to our users, CISER uses vendors both on and off campus to fulfill the technological requirements for the data and reproduction archive. On campus we rely heavily on the expertise provided by the Center for Advanced Computing (CAC)(4) and Cornell Central IT (CIT)(5). The team at CAC provides file storage space for archive data, infrastructure as a service (IAAS) and maintains our fleet of high-performance research servers utilized by students and researchers alike across non-restricted and restricted (Cornell Restricted Access Data Center) environments. The new CISER Data and Reproduction Archive hosts all of our files on Cornell Campus through CAC's file storage space and runs the download server from a virtual server on CAC's IAAS platform. The file server is backed up daily by CIT's "EZ-Backup Service". This is a full snapshot of our data holdings.

Externally, Cornell University through Cornell IT has entered into an enterprise agreement with Amazon Web Services (AWS). Access to this service requires yearly attestations to be signed along with following all prescribed security and logging requirements set forth by the Cloudification Team. The new CISER Data and Reproduction Archive hosts the frontend UI, API, and Metadata database in AWS. Backups of the UI and API are made daily and are geographically distributed to ensure no loss of data should a disaster strike an individual data center, although both can be recreated instantly from code hosted in a GitHub repository. The database is hosted in a multi-availability zone configuration meaning an entire AWS data center could go down without data-loss. Backups are also made daily and are kept in a rolling 30 day window.

In addition, CISER uses tools that have been developed outside to enhance service delivery, such as StatTransfer for data conversion, and Trello and OpenProject for project tracking purposes. CISER maintains formal relationships with the Interuniversity Consortium for Political and Social Research (ICPSR), Data-PASS, the Roper Center, and the Data Documentation Initiative Alliance.

Reviewer Entry**Reviewer 1**

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:
Accept

Summary of Significant Changes Since Last Application (if applicable).

We ended our Dataverse instance and rebuilt our internal archive database for improved preservation and discoverability.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:
Accept

Other Relevant Information.

See CISER's Re3data record here; <https://www.re3data.org/repository/r3d100011056>

CISER is a partner center in the new Cornell Center for Social Sciences that is housed within Cornell's Office of the Vice Provost for Research (OVPR), and 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.(6) 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.(7) At present, the archive contains nearly 2,000 studies and 22 R-Squared packages with approximately 22,300 individual files totaling about 713 gigabytes plus four data download centers. In addition, CISER holds a collection of more than 650 studies on CD and DVD.

Links to Supporting Documentation: (all links visited 2020-09-29)

1. About CISER data: <http://ciser.cornell.edu/data/data-archive/>
2. "The FAIR Data Principles," Force11. <https://www.force11.org/group/fairgroup/fairprinciples>
3. "The DCN Curation Workflow," Data Curation Network. <https://datacurationnetwork.org/resources/workflows/>
4. Center for Advanced Computing: <https://www.cac.cornell.edu/>
5. IT@Cornell: <https://it.cornell.edu/>
6. Cornell University Research Division: <https://research.cornell.edu/research-division>
7. About CISER – Partners: <https://ciser.cornell.edu/about-us/partners/>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:
Accept

ORGANIZATIONAL INFRASTRUCTURE

1. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:
4 – The guideline has been fully implemented in the repository

Response:

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 data needs of Cornell social scientists throughout the entire research process and data life cycle is integrated into organizational policy, procedure, and practice.

The CISER Data and Reproduction Archive holds an extensive collection of data studies and reproduction packages in the social sciences, with emphasis on demography, economics and labor, political and social behavior, family life, and

health. Long-term preservation is a key component of our Collection Development Policy(3) and core to our Data and Reproduction Archive's mission.(4)

CISER provides consulting services to identify, obtain, and use datasets as well as data management support throughout the research lifecycle. CISER has fully trained staff who work with researchers and data providers to ensure that data and accompanying documentation comply with CISER standards and policies.(3)(5)(6) CISER also provides instruction in the use of a variety of statistical software packages as well as individual troubleshooting consultations.(7)

CISER develops, monitors, and updates its mission statement through an internal process involving all relevant 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. Effective July 1, 2020, CISER became a partner center within the new Cornell Center for Social Sciences (CCSS).(8) Changes to the mission statement are reviewed and finalized with the CCSS and shared with member colleges.

Links to Supporting Documentation: (all links visited 2020-09-29)

1. CISER history: <https://ciser.cornell.edu/about-us/history/>
2. CISER Mission and Vision: <https://ciser.cornell.edu/about-us/strategic-plan/>
3. CISER Data and Reproduction Archive Collection Policy:
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-collection-policy/>
4. Data and Reproduction Archive Mission Statement:
<https://ciser.cornell.edu/data/data-archive/data-archive-mission-statement/>
5. CISER Data and Reproduction Archive Preservation and Storage Policy:
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-preservation-and-storage-policy/>
6. CISER Resources Terms of Use: <https://ciser.cornell.edu/about-us/ciser-policies/ciser-resources-terms-of-use/>
7. CISER Training: <https://ciser.cornell.edu/training/>
8. Cornell Center for Social Sciences: <https://www.socialsciences.cornell.edu/>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:
Accept

2. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:

4 – The guideline has been fully implemented in the repository

Response:

For all datasets CISER complies with data producer terms and conditions through data use agreements signed by the Data Librarian. In addition, CISER has annual memberships with ICPSR, 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 users must observe the CISER Resources Terms of Use policy(1) and agree to the CISER Data and Reproduction Archive Acceptable Use Policy, which is displayed to the researcher as they download files.(2) CISER data 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.(3)(4) In addition to restriction computing access if users violates our policies, CISER reserves the right to demand the return of data and deny future access to data if a user is in violation of data use agreements, CISER's policies, or Cornell University policies.

The dissemination of data from the CISER Data and Reproduction Archive is built upon a six-level restriction system to signify whether data is publicly available, restricted to Cornell researchers, or, based on data producer terms and conditions, requires approval outside of the normal download process.(5) Users are responsible for maintaining that they are in compliance with security safeguards required for the type of data. They must also agree to adhere to licensing requirements as stipulated by the data provider and presented upon download.

Links to Supporting Documentation: (all links visited 2020-10-28)

1. CISER Resources Terms of Use Policy - <https://ciser.cornell.edu/about-us/ciser-policies/ciser-resources-terms-of-use/>

2. CISER Data and Reproduction Archive Acceptable Use Policy -

<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-acceptable-use-policy/>

3. Cornell University Policy Office (Policy Volume V. Information Technologies)

<https://www.dfa.cornell.edu/policy/policies/responsible-use-information-technology-resources>

4. Cornell University IT Policies - <https://it.cornell.edu/policy>

5. Restriction Legend: <https://archive.ciser.cornell.edu/about/file-levels>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Accept.

3. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The fundamental purpose of CISER's Data and Reproduction 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 and Reproduction Preservation and Storage Policy documents the main theoretical and practical steps for providing long-term preservation of digital research data.(1) See Section X for a description and discussion of CISER's preservation policy. CISER routinely monitors technical developments, as described in Sections X and XIV of this document.(1, Management of Storage Infrastructure)

CISER operates within the framework of the Cornell University Continuity of Operations plan.(2) 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.(3)

Long-term data continuity is also provided by the Data Preservation Alliance for the Social Sciences (Data-PASS), a partnership to archive, catalog, and preserve data used for social science research of which CISER is a member and has a Memorandum of Understanding.(4) This MOU states that "In the event that one of the partner institutions should no longer be able to retain the material they acquire, and no succession plan is already in place, the partner's materials will be transferred to one of the other partners to meet the objectives of the partnership."(5)

As mentioned elsewhere in this document, off-site backup of the archive is performed on a daily basis.

Links to Supporting Documentation: (all links visited 2020-03-30)

1. CISER Data and Reproduction Archive Preservation and Storage Policy:

<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-preservation-and-storage-policy/>

2. Cornell University Continuity of Operations Plan:

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

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

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

5. Sample MOU available: http://data-pass.org/sites/default/files/Data-PASS_MoU_201504.pdf

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Accept.

4. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:

4 – The guideline has been fully implemented in the repository

Response:

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 reviews 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. (1, "Purchasing Data") In the case of CISER's Results Reproduction Service (R-squared), CISER staff reproduce the findings to ensure the publication is accurately representative of the data.(2)

As mentioned previously, the majority of CISER's data files are public use files that are available in the CISER Data and Reproduction Archive for widespread access. Forty to fifty percent of the collection are only available to the Cornell community and a very small subset of those require application approvals for access. The data are either licensed to the Cornell community only or in some cases a researcher has asked for the ability to approve access. Per the Collective Development Policy, CISER Data and Reproduction Archive will not accept data which contains Personally Identifiable Information (PII) except in such cases where the data are part of the public record.(1)

Any users of Cornell networks must abide by Cornell University policies which are in line with generally accepted higher education policies.(3, 4) Any users downloading data from the CISER Data and Reproduction Archive must agree to the CISER Acceptable Use Policy prior to downloading datasets.(5) 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.(6, 7)

In addition to agreeing to the Acceptable Use Policy, if a user accesses data through our computing accounts, CISER reserves the right to disable a computing account immediately upon identification of possible misuse of any 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 2020-10-20)

1. CISER Data and Reproduction Archive Collection Policy:

<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-collection-policy/>

2. CISER: Results Reproduction Service: <https://ciser.cornell.edu/research/results-reproduction-r-squared-service/>

3. Cornell University Policy Office: <https://www.dfa.cornell.edu/policy>

4. Cornell University Campus Code of Conduct: <https://www.dfa.cornell.edu/policy/policies/campus-code-conduct>

5. CISER Data and Reproduction Archive Acceptable Use Policy:

<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-acceptable-use-policy/>

6. CISER Computing Accounts: <https://ciser.cornell.edu/computing/>

7. CISER Resources Terms of Use: <https://ciser.cornell.edu/about-us/ciser-policies/ciser-resources-terms-of-use/>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Accept.

5. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:

4 – The guideline has been fully implemented in the repository

Response:

CISER is legally considered part of Cornell University and is a Partner Center of the Cornell Center for Social Sciences which is housed within the Office of the Vice Provost for Research.(1) CISER is funded through Cornell University on a structural basis. The university operates on an annual budget model, although it is planned out for multiple years. Based on the ongoing model of funding, as well as Cornell's recent commitment to investing in excellence in the Social Sciences, Cornell has sufficient funding to maintain CISER's services and for attending essential meetings and training for the next 3 – 5 years.(2) CISER currently has 9.35 FTE across 10 staff, with 2 FTE archive staff and another .5 FTE of CISER staff dedicated to archive IT and other support.

Cornell provides training and professional development opportunities that 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 and funds staff attendance. 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 archive. CISER staff hold doctorates in History and Development Sociology and master's degrees in Management, Political Science, Sociology, Development Sociology, History (multiple), and Library and Information Studies. 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, RDAP, the American Sociological Association, SSHA, IUSSP, the American Statistical Association, American Historical Association, Organization of American Historians, American Library Association, and ISSA individually. The Director is past President of IASSIST (4, "Past Presidents") , and currently Executive Director of the Social Science History Association(5) and Vice Chair of the DDI Executive Committee.(6) The Senior Data Librarian is the current Chair of the ALA Government Documents Round Table,(7) the current Secretary of IASSIST,(4) and the Designated Representative to ICPSR. The Research Associate is co-chair of the International Fellows Organization of IASSIST(4, "Committee Chairs") and is Official Representative to ICPSR. The Data Archive Specialist serves as Cornell Organizational Representative to Datacite and on the membership committee of the Research Data Access and Preservation Association (RDAP).(8)

Links to Supporting Documentation: (all links visited 2020-10-28)

1. The Research Division: <https://research.cornell.edu/research-division>

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. DDI Executive Board: <https://ddialliance.org/node/938>
7. GODORT Chair: <http://www.ala.org/rt/godort/chair>
8. RDAP Association Leadership: <https://rdapassociation.org/leadership>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Accept

6. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse or external, including scientific guidance, if relevant).

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:

4 – The guideline has been fully implemented in the repository

Response:

There are multiple mechanisms CISER uses for 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 related to CISER service areas 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. As detailed in Section V, CISER staff are members of many professional organizations, which provide technical, curation, data science, and disciplinary networks for expert advice.

CISER is a founding and active 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 Senior Data Librarian and the CISER Research Associate are consultants for RDMSG.(3) The Data Archive Specialist also supports RDMSG initiatives. RDMSG consultants are available for meetings upon request, via phone, by email, Zoom, or in person (in person is suspended for the academic year 2020-2021 due to COVID-19).

For users, CISER provides a Help Desk service Monday through Friday 8:00 am - 4:00 pm. This service is available online, or by phone, email, or walk-in (walk-in is suspended for the academic year 2020-2021 due to COVID-19).(4) It is fully available for all members of the Cornell community, as well as others who need assistance in using statistical software or preparing code or data. In addition the CISER Senior Data Librarian provides support for data discovery and the CISER Data and Reproduction Archive and other sources through office hours and provides chat support to the Cornell University Library.(5) Her virtual office hours are Wednesdays 2:00 - 3:00 pm and Fridays 10:00 am - 11:00 am during the fall 2020 semester.

To evaluate services, CISER uses TeamDynamix, a ticket tracking system provided by Cornell. After tickets have been resolved, users receive a short survey asking about their satisfaction level with services and suggestions for improvement. In addition, CISER solicits feedback from partners across campus through the RDMSG.

Links to Supporting Documentation: (all links visited 2020-10-28)

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: <https://ciser.cornell.edu/consulting/ciser-helpdesk/>
5. CISER Research Guide: <https://guides.library.cornell.edu/CISER>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:
Accept

DIGITAL OBJECT MANAGEMENT

7. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:
4 – The guideline has been fully implemented in the repository

Response:

DATA INTEGRITY:

A BLAKE3 checksum is generated for every data file added to the CISER Data and Reproduction Archive to ensure integrity of the digital file both now and into the future.(1, "Data Integrity") Checksum validations are run to report all files which have been added, deleted, or modified since the previous validation.(1, "Data Integrity," "Data Normalization")

The determination of completeness of metadata is made at the discretion of the Data Librarian in consultation with the Data Archive Specialist. 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 and Reproduction Archive.(2) CISER assigns a new version to a data study with significant changes such as: addition of new variables; revision of incorrect data; revision of miscoded data; formatting changes; substantial documentation changes; changes in access conditions; withdrawal of data elements or documentation files. Minor changes such as small changes in variable labels, spelling corrections in metadata, minor changes in documentation will be made to relevant content and recorded in ancillary and accompanying documentation, but no new version assigned. Any and all changes are recorded in an internal notes field in our database. CISER follows Data Documentation Initiative (DDI) standards, as well as the Open Archival Information Systems Reference Model, Core Trust Seal, FAIR Data Principles, and Trusted Repositories Audit and Certification.(1) Since the last CoreTrustSeal application, we have scanned physical documentation and added digitized documentation and codebooks to the bulk of the archival collection.

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 Section XI).(2) Direct contact information is available on every page as users navigate through the archive. Users are invited to contact us regarding questions or errors using that email. They may also contact 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 due to significant changes or error, CISER staff enters this information in the file's database record.(2)

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.(3) As described in Section IX, automated tasks are run on a scheduled basis to verify checksums and permissions. The results are compared to the metadata, held within the SQL database, to validate data integrity.(1)

In the new CISER Data and Reproduction Archive, staff 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.(2)

The vast majority of CISER's holdings are purchased or freely available datasets from external providers. While the percentage of Cornell-created datasets remains low (approximately 3% of the collection), this number has risen steadily over the last several years, especially with our R-Squared service. In most cases, Cornell depositors are personally known to the Institute and files are provided to and deposited into the archive by CISER staff. Since the last CoreTrustSeal application, we have instituted a data deposit form in Qualtrics to guide depositors through a series of questions which provides the bulk of the required metadata for a deposit. In addition, the form includes questions to help CISER staff to determine whether the deposit meets the archive's collection policy.(4)

Links to Supporting Documentation: (all links visited 2020-11-17)

1. CISER Data and Reproduction Archive Preservation and Storage Policy:
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-preservation-and-storage-policy/>
2. CISER Data and Reproduction Archive Versioning Policy:
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-versioning-policy/>
3. Browse Data Archive Holdings by Subject: <https://archive.ciser.cornell.edu/explore/subjects>
4. CISER Archive Data Deposit Form: https://cornell.qualtrics.com/jfe/form/SV_eDUAFPGQ9bJUw61

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Accept.

8. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:

4 – The guideline has been fully implemented in the repository

Response:

COLLECTION DEVELOPMENT:

The CISER Data and Reproduction Archive Collection Policy (1) drives data appraisal and acquisition for the CISER research archive. Data acquisition is primarily demand driven. The Data and Reproduction 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.

METADATA:

Upon receipt of new digital content, the Archive staff process the data and documentation in collaboration with the data provider if possible, following a slightly modified version of the Data Curation Network's CURATE(D) process.(2) The metadata pertaining to each data file is stored in a SQL database. Several metadata fields have been pre-filled with our agreed upon controlled vocabularies to ensure adherence to the relevant schemas. For data submitted by a researcher, we will work with the researcher to ensure that the metadata is complete and accurate. 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.

Data are accompanied by relevant 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 files are in line with FAIR principles.(3) The Data and Reproduction Archive reserves the right to reject datasets deemed inadequately documented. (1, "Data NOT in Scope") 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.(4)

FORMATS:

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, encourages use of Library of Congress recommended dataset formats.(1, "Data Curation - File Formats"; 5) 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.

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 and reproduction archive.

CISER staff check submitted datasets for their file formats and contact the depositor, if necessary. (1; 6, “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, “Data Curation - File Formats”)

Links to Supporting Documentation: (all links visited 2020-11-03)

1. CISER Data and Reproduction Archive Collection Policy:

<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-collection-policy/>

2. The DCN Curation Workflow: <https://datacurationnetwork.org/outputs/workflows/>

3. The FAIR Data Principles: <https://www.force11.org/group/fairgroup/fairprinciples>

4. CISER Data and Reproduction Archive Versioning Policy:

<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-versioning-policy/>

5. “Recommended Formats Statement, Datasets” Library of Congress:

<https://www.loc.gov/preservation/resources/rfs/data.html#datasets>

6. CISER Data and Reproduction Archive Preservation and Storage Policy:

<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-preservation-and-storage-policy/>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Accept

9. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The processes and procedures for managing archival storage are documented in the CISER Data and Reproduction Archive Preservation and Storage Policy and internal workflow documentation available on a shared drive.(1) The policies are managed through a coordinated effort between the Senior Data Librarian, the CISER Director, and other CISER staff. Policy changes are approved by the CISER Director. The workflow documentation is managed by the Senior Data Librarian and updated in conjunction with other CISER staff.

Six restriction levels guide the dissemination of data from the CISER Data and Reproduction Archive. The files that are publicly available are designated with green level. The remaining levels are restricted in increasing levels from requiring a Cornell netID to requiring data use agreements or other forms of approval based on provider requirements.(2) Users must abide by the CISER Acceptable Use Policy and any requirements in Data Use Agreements as stipulated by specific data providers.(3) The CISER Acceptable Use Policy and any data use requirements are managed within the archive using click-through instructions. Approval processes are also initiated and managed within the archive's database.

The CISER Data and Reproduction Archive data files are stored on network attached storage (NAS) in Cornell University's Data Center with 187TB capacity. The NAS disk runs RAID 6 and has manufacture call-home features enabled for expedited servicing. 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.

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 and Reproduction Archive. (1, "Management of Storage Infrastructure")

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).(4)

To ensure that the digital content remains identical and accessible across archival copies, automated tasks are run on a scheduled basis to verify checksums. 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 reinstate the original version from a backup copy. After data retrieval, automated tasks 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: (all links visited 2020-11-18)

1. CISER Data and Reproduction Archive Preservation and Storage Policy:

<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-preservation-and-storage-policy/>

2. Restriction Legend: <https://archive.ciser.cornell.edu/about/file-levels>

3. CISER Data and Reproduction Archive Acceptable Use Policy:

<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-acceptable-use-policy/>

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

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Accept.

10. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The CISER Data and Reproduction Archive Preservation and Storage Policy documents the main theoretical and practical steps for providing long-term data preservation. These steps are integrated into archival operations and planning within CISER as part of the research data lifecycle.(1) This 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), Core Trust Seal, Data Documentation Initiative (DDI)), that represent an international body of knowledge and expertise pertaining to various issues within digital preservation.(1) 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.(1)

CISER staff ensure the integrity, completeness, and authenticity of data submitted to the Data and Reproduction Archive during the ingest process as outlined in our Data Collection Policy. CISER has developed a data deposit application to accommodate researchers wanting to deposit their data at CISER.(5) Data transfer methods from the provider to CISER 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 through most transfer methods at the provider's convenience. During the ingest process, non-supported file formats are converted to specified formats that support long-term preservation.(2)

Contracts differ in language and requirements depending on the distributor and method of acquisition. To accommodate the differences in requirements, CISER has six levels for data distribution.(3) Data that cannot be distributed through the data and reproduction archive and restricted data for specific research projects are housed in CISER's Cornell Restricted Access Data Center (CRADC) and requirements are tracked with Agiloft software.(4) CRADC is a separate entity from the CISER Data and Reproduction Archive at this time. 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 archive and access is restricted until the embargo is over.

Measures to ensure these actions are taken currently reside with the Data Librarian. CISER routinely monitors technical developments (standards, software, tools, and platforms) and evaluates potential archival solutions that will both streamline and enhance data preservation and archival practices.(1, "Management of Storage Infrastructure")

Links to Supporting Documentation: (all links visited 2020-11-19)

1. CISER Data and Reproduction Archive Preservation and Storage Policy:

<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-preservation-and-storage-policy/>

2. CISER Data and Reproduction Archive Collection Policy:

<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-collection-policy/>

3. Restriction Levels: <https://archive.ciser.cornell.edu/about/file-levels>

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

5. CISER Data and Reproduction Archive Data Deposit Form:
https://cornell.qualtrics.com/jfe/form/SV_eDUAFPGQ9bJUw61

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:
Accept

11. Data quality

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Data producers seeking to deposit data in the CISER Data and Reproduction Archive must provide metadata in compliance with domain standards. Data studies should be accompanied by comprehensive and relevant documentation:

codebooks, file layout maps, technical notes, survey instruments questionnaires, reports, and errata in open and accessible formats.(1) As mentioned elsewhere, when possible, our Senior Data Librarian and Data Archive Specialist work with data providers to provide clear and complete metadata. Both the Data Librarian and the Data Archive Specialist have had training in the Data Curation Network’s CURATED method, DDI, and years of experience working with data and metadata. (See Section VII for description of the data deposit process.) CISER adheres to DDI Lifecycle standards for metadata(2) and the FAIR data principles.(3) The SQL database is mappable to DDI at the study level.

CISER accepts feedback on all user issues including data and metadata quality through email or the help desk using a TeamDynamix ticketing system, and informally at meetings and conferences.(4)(5) 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 6 for an example) For reproduction packages, citations to related journal articles are always provided. Acknowledgement of use of CISER support and resources is requested with a provided acknowledgement statement.(7) In addition, the CISER Data and Reproduction Archive provides a suggested citation for each data study and reproduction package.

Links to Supporting Documentation: (all links visited 2020-10-09)

1. CISER Data and Reproduction Archive Collection Policy:
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-collection-policy/>
2. CISER Data and Reproduction Archive Preservation and Storage Policy:
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-preservation-and-storage-policy/>
3. The FAIR Data Principles: <https://www.force11.org/group/fairgroup/fairprinciples>
4. CISER Help Desk Services: <http://ciser.cornell.edu/consulting/ciser-helpdesk/>
5. TeamDynamix: <https://it.cornell.edu/teamdynamix>
6. How Parents Fare: Mothers’ and Fathers’ Subjective Well-Being in Time with Children:
<https://archive.ciser.cornell.edu/reproduction-packages/2775/related-articles>
7. Acknowledgement: <https://ciser.cornell.edu/acknowledgement/>

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Reviewer 3

Comments:

Accept

12. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:

4 – The guideline has been fully implemented in the repository

Response:

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 and Reproduction Archive catalog;
- data preservation (data integrity, normalization, storage infrastructures) (1)

CISER staff has comprehensive documentation of policies available publicly (1) and documented workflow procedures available internally. Data studies are appraised for possible ingest based on CISER's Data and Reproduction Archive Collection Development Policy.(2) If data are not within the scope of CISER's Archive, efforts are made to suggest possible alternative archives.

The CISER Data and Reproduction Collection Policy (2) details criteria and information regarding the selection of data for archiving. Data acquisition is primarily demand driven. Archive staff will attempt to acquire any data required by faculty members in accordance with organizational policies regarding cost, quality, restrictions, and expected future use. Archive staff will not accept data which contains personally identifiable information, 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 the Cornell Restricted Access Data Center (CRADC)(3) provides secure access. (2, "Data Not in Scope")

In cases where documentation is insufficient, CISER works with data producers to ensure that data files are usable and

understandable by generating additional contextual information. 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 migrate data into these formats and retain the original copies.(2)

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

CISER staff who manage data follow a modified version of the Data Curation Network's CURATED workflow that they adhere to while documenting ingest processes and data transformations.(4) Other processes such as long-term preservation (e.g. normalization, version control, sustainability) are detailed in the CISER Data and Reproduction Preservation and Storage Policy(5) and CISER Data and Reproduction Archive Versioning Policy.(6)

Workflows for packages created by the Results Reproduction (R-Squared) service are maintained using OpenProject software. Workflows for data studies requiring Data Use Agreements are maintained separately for each study within the archive database to ensure that the proper requirements for dissemination are met.

User direction is provided in the Archive's Acceptable Use Policy (7) which must be agreed to before files can be downloaded. These directions indicate that downloaded resources are for statistical analysis only, require protection of confidentiality of all research participants, and to remain in accordance with any and all Data Use Agreements and licensing requirements as stipulated by the providers of datasets. This information is shown in a statement whenever the user requests to download a file.

Additional permission settings can be created on the server for studies which require authorization from the data provider. We have several levels in our archive to accommodate additional provider requirements, from a simple attestation on file, to a required electronic signature, to extended or mediated approval processes.(8) We are able to conduct most functions through the database's web interface directly and easily.

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. (5, "Management of Storage Infrastructure")

Links to Supporting Documentation: (all links visited 2020-11-17)

1. CISER Policies (Universal and Data and Reproduction Archive): <https://ciser.cornell.edu/about-us/ciser-policies/>
2. CISER Data and Reproduction Archive Collection Policy:
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-collection-policy/>
3. CRADC: <https://ciser.cornell.edu/data/secure-data-services/cradc/>
4. The DCN Curation Workflow: <https://datacurationnetwork.org/resources/workflows/>

5. CISER Data and Reproduction Archive Preservation and Storage Policy:
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-preservation-and-storage-policy/>
6. CISER Data and Reproduction Archive Versioning Policy:
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-versioning-policy/>
7. CISER Data and Reproduction Archive Acceptable Use Policy:
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-acceptable-use-policy/>
8. Restriction Legend: <https://archive.ciser.cornell.edu/about/file-levels>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:
Accept

13. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Search facilities are available online at: <https://archive.ciser.cornell.edu/search>

The CISER Data and Reproduction Archive's online catalog is listed on the Registry of Research Data Repositories (re3data.org) (1) and offers robust search facilities to enable discovery of and access to data files, codebooks and other documentation materials. Users can search for data by conducting free text searching with priority given to title, keywords, and abstract. The results can be filtered by subject, principal investigators, producers, grant agencies, collection methods, series, and whether the studies are R-Squared packages, Cornell authored, or enhanced by CISER. The catalog can also be browsed by subject, series, new additions, and results reproduced status. (2 (particularly "Data Normalization"), 3, 4) The catalog record includes a suggested citation, which can be copied to the researcher's clipboard with a single click.

The Data and Reproduction Archive's catalog is preserved by migrating the collection to new versions or when new formats become widely available. 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), Core Trust Seal CTS), and Data Documentation Initiative (DDI)). (2, "Reason for Policy")

All data studies maintained by the CISER Data and Reproduction Archive are assigned a locally-generated unique identifier see <https://archive.ciser.cornell.edu/reproduction-packages/2775> for example, where 2775 serves as the unique identifier). CISER has assigned Digital Object Identifier persistent IDs to all new data records and to most of the legacy files in the archive. Since our last CoreTrustSeal application the Archive has transitioned from EZ-ID to full membership in Datacite and serves as the lead repository for Cornell University.(5)

The CISER Data and Reproduction Archive does not currently have an API to programmatically grab metadata. However, the archive's data studies and reproduction package pages contain embedded JSON-LD markup. This allows compatible tools such as Google Dataset Search to crawl and index our holdings. We have plans to add an API in a future iteration of the archive.

Links to Supporting Documentation: (all links visited 2020-11-17)

1. Re3data.org record for CISER: <http://www.re3data.org/repository/r3d100011056>
2. CISER Data and Reproduction Archive Preservation and Storage Policy: <https://ciser.cornell.edu/about-us/ciser-policies/data-archive-preservation-and-storage-policy/>
3. CISER Data and Reproduction Archive: Online catalog: <https://archive.ciser.cornell.edu/>
4. Browse Data and Reproduction Archive Holdings: <https://archive.ciser.cornell.edu/explore>
5. DataCite Members (See Cornell University): <https://datacite.org/members.html>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:
Accept

14. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:
4 – The guideline has been fully implemented in the repository

Response:

METADATA:

Data producers seeking to deposit data in the CISER Data and Reproduction Archive must provide metadata in compliance with domain standards (DDI, OAIS, etc.). For researchers, CISER staff provide a data deposit form asking for required information including questions about responsible parties, title, description, methodology, and additional information. Metadata is assessed to ensure it meets minimum requirements, and CISER staff work with researchers and data providers to develop comprehensive metadata.(1) Data studies and reproduction packages should be accompanied by comprehensive documentation: codebooks, file layout maps, technical notes, questionnaires, reports, and errata in open and accessible formats. (2, "Data Curation") In addition, the CISER Archive's fields are mapped to DDI standards.

FILE FORMATS:

CISER prefers file formats in the Library of Congress list of recommended formats.(3) Data are provided primarily in csv or

ASCII format. In addition, we are converting older ASCII files to csv formats. Some datasets include SAS, SPSS, Stata, and/or R-ready files.(4, “Data Normalization”) During the ingest process, non-supported file formats are converted to supported formats. Evaluation of new content types and software/format obsolescence is an ongoing process.

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.(4 (“Reason for Policy,” Data Normalization”), 5) Available documentation is stored on-site and in digital formats. CISER staff have converted most physical documentation to digital format for older studies and will continue to convert as necessary.. A number of software packages (SAS, SPSS, Stata, R, Matlab, etc.) are also provided within our computing environment to make data more accessible to researchers.(6)

CISER routinely monitors technical developments (standards, software, tools, and platforms) and evaluates potential archival solutions that will both streamline and enhance data preservation and archival practices. (4, “Management of Storage Infrastructure”)

Links to Supporting Documentation: (all links visited 2020-11-20)

1. CISER Data and Reproduction Archive Deposit Form: https://cornell.qualtrics.com/jfe/form/SV_eDUAFPGQ9bJUw61
2. CISER Data and Reproduction Archive Collection Policy :
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-collection-policy/>
3. Library of Congress Recommended Formats Statement: <https://www.loc.gov/preservation/resources/rfs/TOC.html>
4. CISER Data and Reproduction Archive Preservation and Storage Policy:
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-preservation-and-storage-policy/>
5. CISER Data and Reproduction Archive Versioning Policy:
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-versioning-policy/>
6. Available Software: <https://ciser.cornell.edu/computing/computing-account-software/>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:
Accept

TECHNOLOGY

15. Technical infrastructure

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:

4 – The guideline has been fully implemented in the repository

Response:

CISER evaluated several high-profile projects within the domain of archival work and came to the determination that “off-the-shelf solutions” would not meet the specific needs of our data and reproduction archive. Therefore, CISER elected to create software in house in order to satisfy specific needs. The frontend software is written in TypeScript / ReactJS. All backend services including authentication / authorization, archive API, DOI resolver and download generation server are implemented in Java. Upcoming supplemental systems as the system grows will likely target Java, but this will be evaluated based on the task desiring to be accomplished. All servers are running the most recent, long-term support release of Ubuntu with all patches installed.

The relational database is hosted on an Amazon Web Services (AWS) PostgreSQL Relational Database Server (RDS) instance. This database contains our schema which is mappable to DDI, Dublin Core and other leading standards, along with metadata and permissions information for our holdings that consist of studies, reproduction packages and download centers. All new holdings feature a DOI (hosted by DateCite) and almost all legacy holdings have been issued a DOI.(1)

Our technical team composition requires a hybrid approach to hosting the archive. Frontend technologies, the API, the DOI resolver and relational database are all hosted in the cloud on AWS services. Authentication / Authorization and download servers are hosted within Cornell University's private cloud environment Red Cloud. Archive files are stored within CISER's private network. This separation allows for our team to develop rapidly while conforming with data agreements on where materials are hosted.

All parts of the archive fall under a disaster recovery plan or can simply be redeployed effortlessly. The RDS instance is hosted within a multi-availability zone and is backed up daily. These backups are kept in rolling storage for 30 days, allowing for near instant rollback should an incident arise. Our file storage capability on campus is covered by Cornell Information Technology EZ-Backup service daily. This service provides a two-week rolling backup window. All files within the archive adhere to CISER's Data Preservation and Storage Policy.(2)

We have sufficient availability and bandwidth for this operation. Across all of the services utilized, AWS provides a minimum Service Level Agreement (SLA) monthly uptime percentage of 99.9%. CISER's 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).(3)

Links to Supporting Documentation: (all links visited 2020-11-20)

1. Datacite: <https://www.datacite.org/>
2. CISER Data and Reproduction Archive Preservation and Storage Policy: <https://ciser.cornell.edu/about-us/ciser-policies/data-archive-preservation-and-storage-policy/>
3. Cornell Wired Network: Faculty and Staff - <https://it.cornell.edu/wired>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:
Accept

16. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Reviewer 3

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Archive data is backed up daily by Cornell Information Technology EZ-Backup service (using Tivoli Storage Manager [TSM]). Metadata is backed up daily via Amazon Web Services RDS. These backups are held for a rolling 30-day period. Additionally, our PostgreSQL database is hosted in a multi-availability zone configuration to help mitigate both downtime and data loss due to a failure in a datacenter. Each archive file has a cryptographic hash checksum (BLAKE3) that is stored alongside other file metadata. A scheduled task runs daily verifying that all archive files match their recorded checksums.

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 replication. 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 and Reproduction 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)

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 administrative interface is controlled by authentication through Shibboleth (using Cornell Central IT authentication) and is authorized by permissions built into the purpose-built archive system. End users also authenticate with Shibboleth in order to download files that require one to be a member of the Cornell community or those that require additional granted permissions.

Access to the file store where the archive files are hosted is secured by NTFS file permissions and active directory roles. Only authorized CISER staff have accounts on the CISER Research Network that have read / write access to this location. These accounts are completely isolated from the standard Cornell University NetID account. Any original media/electronic data that is retained, will be stored in compliance with the CISER Data and Reproduction Archive Preservation and Storage Policy. (<https://www.dfa.cornell.edu/policy/policies/reporting-electronic-security-incidents1>) Reporting security incidents is mandated by Cornell University Policy 5.4.2, Reporting Electronic Security Incidents. (4)

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

The CISER Software Engineer is the local security personnel. In addition, CISER contracts with the Cornell Center for Advanced Computing for System Administration supported by the Assistant Director of Systems. (6)(7)

Links to Supporting Documentation: (all links visited 2020-11-20)

1. CISER Data and Reproduction Archive Preservation and Storage Policy:
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-preservation-and-storage-policy/>
2. CISER Data and Reproduction Archive Security Policy:
<https://ciser.cornell.edu/about-us/ciser-policies/data-archive-security-policy/>
3. Cornell University Policy 8.4 – Management of Keys and Other Access Control Units (see pp. 8-18 primarily) -
<https://www.dfa.cornell.edu/policy/policies/management-keys-and-other-access-control-systems>
4. Cornell University Policy 5.4.2 – Reporting Electronic Security Incidents: <https://www.dfa.cornell.edu/policy/policies/reporting-electronic-security-incidents>
5. Solarwinds Log and 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

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:
Accept

APPLICANT FEEDBACK

Comments/feedback

These Requirements are not seen as final, and we value your input to improve the CoreTrustSeal certification procedure. Any comments on the quality of the Requirements, their relevance to your organization, or any other contribution, will be considered as part of future iterations.

Response:

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Reviewer 3

Comments:

All comments have been addressed. I recommend accepting the application.