ScholarsArchive@OSU

Notes Before Completing the Application

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

True

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:

Institutional repository, Publication repository, Library/Museum/Archives

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Brief Description of Repository

ScholarsArchive@OSU <https://ir.library.oregonstate.edu/> is Oregon State University's (OSU) institutional repository for the research, scholarship, and publications of the university. The Emerging Technologies and Services Department at OSU Libraries and Press is responsible for maintaining the technological infrastructure for the repository. Repository holdings include all of the university's doctoral, masters, and undergraduate level theses and dissertations, and masters level final research projects; a significant percentage of affiliated faculty and student research articles; research datasets; conference papers and proceedings; and other content that supports the teaching and research mission of the university. In this application, we discuss both types of research--research datasets and textual documents--contained within the repository because we believe that all of this content is within the scope of Core Trust Seal certification.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Brief Description of the Repository's Designated Community.

The vast majority of items in ScholarsArchive@OSU are public and can be freely accessed by anyone in the world with an internet connection. Users do not need to register or login to view or download items. A small percentage of items are available only to members of the OSU community at the discretion of the author.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Level of Curation Performed. Select all relevant types from:

B. Basic curation – e.g. brief checking; addition of basic metadata or documentation, C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Comments

After deposit and prior to publication, all content is reviewed by administrators to ensure that it is associated, using metadata, with appropriate collections and is able to be downloaded. Administrators review and enhance metadata for all content contributed to the repository except research articles deposited in compliance with the Faculty Senate approved Oregon State University Open Access Policy <https://library.oregonstate.edu/open-access>. These are made available without undergoing review.

Data Curators review all research dataset deposits and associated metadata to ensure that this content is discoverable, downloadable, understandable, usable, and shareable according to the license selected by the author. Data Curators provide instruction <https://guides.library.oregonstate.edu/Scholars-Archive/Datasets> to depositors about file preparation, preferred file formats, documentation, and licensing. A documentation file (ReadMe file) is mandatory for all dataset deposits and reviewed and approved by a Data Curator prior to publication. A ReadMe template is available to researchers: <https://guides.library.oregonstate.edu/ld.php?content_id=45294345>.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Outsource Partners. If applicable, please list them.

ScholarsArchive@OSU uses Amazon Web Services (AWS) Simple Storage Service (S3) <https://aws.amazon.com/s3/> for backups. MetaArchive <http://metaarchive.org> is used for the selective replication of repository content that is deemed to be most at risk and of the greatest historical importance to the university: the university's corpus of Electronic Theses and Dissertations and Extension and Experiment Station Communication publications. MetaArchive is not used for storing any other repository content and is a dark archive.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

Other Relevant Information.

Since 2005, ScholarsArchive@OSU has served as the Oregon State University institutional repository with a primary goal of providing freely-available and long-term access to the university's historical and contemporary intellectual work, as well as actively preserving these works. ScholarsArchive@OSU also includes materials from outside the institution supporting the university's land, sun, sea, and space grant missions and other research interests. For example, the repository hosts natural resources research from across the Greater Northwest region of the United States for indexing, search, and retrieval in the Oregon Explorer natural resources digital library <https://oregonexplorer.info>, a collaboration between OSULP and the Oregon Institute of Natural Resources.

The repository manages over 74,000 individual files totaling over 3.0 TB, expressed in 18 unique MIME types. As of October 2021, ScholarsArchive@OSU hosts 132 datasets. OSULP strives to maintain the repository in an open and technologically neutral manner to take advantage of evolving digital infrastructure and an international community of open source repository software developers.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

ORGANIZATIONAL INFRASTRUCTURE

I. 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:

First approved in 2015 and revised in 2019 and 2022, the OSU Libraries Digital Preservation Policy
<https://wiki.library.oregonstate.edu/confluence/display/RP/OSU+Libraries+Digital+Preservation+Policy>, describes the
Oregon State University Libraries' mission "to make accessible and hold in trust for future use" the datasets and other
digital resources housed in the ScholarsArchive@OSU repository. The policy reflects values expressed in the Libraries'
current Strategic Plan <https://library.oregonstate.edu/sites/default/files/osulp_strategic_plan_2018-2023.pdf>, particularly
Goal 4 of the strategic plan, which is to "practice active, respectful stewardship of the cultural and intellectual output of our
communities." The OSU Libraries Digital Preservation Policy was approved by Oregon State University Libraries'
Administration, Management and Planning group (LAMP). LAMP is responsible for addressing broad Libraries and Press
initiatives such as strategic planning and library policies.

The data curation activities of OSU Libraries and Press result from our mission to gather, index, and make available
intellectual content of the Oregon State University community through ScholarsArchive@OSU. The primary purpose of

hosting data in ScholarsArchive@OSU is to facilitate data sharing for the purposes of data reuse and contributing to open science, and to preserve the data for this reuse. ScholarsArchive@OSU has a Preservation Policy <https://wiki.library.ore gonstate.edu/confluence/display/RP/ScholarsArchive@OSU+Policies#ScholarsArchive@OSUPolicies-PreservationPolicy >, revised in May 2022, that describes the role of the repository as "both a preservation and an access repository."

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

II. 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:

All depositors to ScholarsArchive@OSU must agree to grant Oregon State University a non-exclusive distribution license according to the ScholarsArchive@OSU Deposit Agreement <https://ir.library.oregonstate.edu/agreement>. The visibility level of a work in ScholarsArchive@OSU determines who is able to access that work. The visibility level for a work is determined by either the access control rules for the collection in which the content is a part, or by the depositor during the

submission process. Visibility options are explained in the Deposit Agreement, and include (1) Public--anonymous public access and use; (2) Oregon State University--restricted access by identified individuals authenticated via university institutional credentials; or (3) Private--only the metadata and abstract for the deposit is accessible. Depositors may also choose to apply an embargo to further control access over their work for a period of time, as described in the ScholarsArchive@OSU User Guide <https://guides.library.oregonstate.edu/Scholars-Archive/Embargoes>. The process for selecting an embargo differs based on the type of document that is uploaded. For example, setting an embargo on a research dataset requires only that the depositor select the embargo they wish to use, while setting an embargo on a Graduate Thesis requires the approval of the Graduate School.

Depositors of research datasets are required to select an appropriate Creative Commons or software license for their dataset. A copyright statement chosen from the options at RightsStatements.org is required for every work in ScholarsArchive@OSU. Licenses and rights statements are captured and stored as part of the metadata for the repository record. Additional guidance for depositors selecting a rights statement and license is provided in the ScholarsArchive@OSU User Guide <https://guides.library.oregonstate.edu/Scholars-Archive/Copyright>.

The Terms of Use for ScholarsArchive@OSU <https://ir.library.oregonstate.edu/terms> state that repository content may be used in any manner not prohibited by copyright or other applicable law. The permissions granted to users of ScholarsArchive@OSU terminate automatically upon any breach by the user of the Terms of Use. OSU does not actively monitor usage to identify instances of noncompliance.

Takedown notices related to copyright or intellectual property concerns submitted to SA@OSU will be referred to OSU's designated Digital Millennium Copyright Act Agent.

Unrestricted, sensitive, and confidential data are defined by OSU's Office of Information Security in the University Data Management, Classification, and Incident Response Policy <https://policy.oregonstate.edu/UPSM/08-015_university_data_management_policy>. Of these, only unrestricted data - "data intended for appropriate general use within the university" - may be eligible for deposit to ScholarsArchive@OSU. Confidential or sensitive data are not accepted in ScholarsArchive@OSU in any form. Unrestricted, sensitive, and confidential data correspond to levels 1, 2, and 3, respectively, as defined by OSU's Human Research Protection Program and Institutional Review Board: <https://research.oregonstate.edu/irb/policies-and-guidance-investigators/guidance/data-security>. Researchers interested in depositing unrestricted human subjects data must communicate that the dataset has human subjects data during the submission process. They will work with a Data Curator; and, when needed, OSU's Office of Research Integrity to ensure that the dataset complies with all the regulations, laws, and ethical standards applicable to the dataset.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

III. Continuity of access

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

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase
Accepted as amended.

Reviewer 2

Comments:

Reviewer 3

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

Response:

Oregon State University <http://www.oregonstate.edu> is Oregon's largest comprehensive public research university, preeminent for both scholarly achievement and the direct impact of applied development. The university celebrated its sesquicentennial in 2018 and in 2021 celebrated research funding that topped $380 million for the fifth straight fiscal year. OSU holds a top-tier research designation from the Carnegie Foundation. The OSU Research Agenda <https://research.oregonstate.edu/research-agenda>, integrated with the university's Strategic Plan 4.0 <https://leadership.oregonstate.edu/strategic-plan>, guides faculty inquiry in OSU's three Signature Areas of Distinction: Advancing the Science of Sustainable Earth Ecosystems; Improving Human Health and Wellness; Promoting Economic Growth and Social Progress. Oregon State is one of only two land, sea, space, and sun grant institutions in the U.S., and is committed in its teaching, research, and public service to fulfilling the land-grant mission to serve the public good.

The ScholarsArchive@OSU repository is seen by the University as a significant contributor to this mission, housing every thesis and dissertation ever produced at the university (over 30,000 items), the full breadth of the Extension and Experiment Station Publications Office publications, in addition to an increasing percentage of faculty and student research articles (over 50% as of 2015--https://doi.org/10.7710/2162-3309.1208), and increasing numbers of research

datasets associated with these publications. The majority of datasets are associated with University theses and faculty research articles. Although any dataset created at OSU is in scope, ScholarsArchive@OSU primarily curates and preserves datasets that are not suited for more appropriate, discipline-specific repositories or for which the creator prefers local curation and preservation. OSU Libraries works with contributors to identify other trusted data storage and repository options as appropriate

<https://guides.library.oregonstate.edu/research-data-services/data-management-archive-preserve>. Another intent of the repository is to support those researchers (including students) who are unfamiliar with the data sharing process and need guidance.

ScholarsArchive@OSU is committed to permanent hosting and preservation of all datasets submitted to the repository. In the unlikely event that OSU and OSULP are unwilling or unable to continue to offer ScholarsArchive@OSU as a service to the University community, it will work with content contributors and supervising curators to identify other curatorial institutions, within or outside the OSU system, willing to take on future custodial responsibility for the content.

The OSU Libraries Research Data Services group provides guidance and support to members of the OSU community for all aspects of the data lifecycle, from planning data management strategies during the grant proposal phase through preserving data in ScholarsArchive@OSU or other venues at the conclusion of projects. These services are free of charge to members of the OSU community. ScholarsArchive@OSU takes care to ensure that those research datasets that are deposited to the repository are usable to researchers who may discover, download, and use the data. To this end, the deposit of all research data is reviewed by a Data Curator. Data Curators ensure that datasets are described with adequate metadata, that file names are descriptive and understandable, and that the data is in a file format appropriate for long term preservation. The ScholarsArchive@OSU Preferred File Formats document is provided to assist depositors with understanding and selecting appropriate formats. <https://guides.library.oregonstate.edu/c.php?g=708682&p=8342195>. If the data is recorded in formats that the Data Curator is not familiar with, the depositor and the reviewer discuss the requirements for data formats, and evaluate alternatives. Research data deposit instructions are available in the Datasets User Guide <https://guides.library.oregonstate.edu/Scholars-Archive/Datasets>. Deposits are available at ScholarsArchive@OSU only after datasets and the accompanying documentation have been approved by a Data Curator.

To date, ScholarsArchive@OSU has been able to host all datasets submitted by OSU researchers, including a few large datasets of more than 1TB. In the case that ScholarsArchive@OSU won't be able to accept exceptionally large datasets created by OSU in the future, curators will work with researchers to identify other appropriate repositories.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Reviewer 3

Comments:

IV. 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

Response:

Metadata staff within the Emerging Technologies and Services department responsible for managing the repository, are responsible for reviewing, assessing, and enhancing metadata associated with all items in the repository. Data Curators, working closely with the staff of the Emerging Technologies and Services department, have primary responsibility for coordinating and reviewing the deposit of datasets into ScholarsArchive@OSU. ScholarsArchive@OSU administrators remove inappropriate data as it is recognized or upon request according to the Research Data Curation Policy <https://wiki.library.oregonstate.edu/confluence/display/RP/ScholarsArchive@OSU+Policies#ScholarsArchive@OSUPolicies-ResearchDataCurationPolicy>.

Repository staff refer people to other appropriate repositories for things that don't fall within the ScholarsArchive@OSU Collection Policy <https://wiki.library.oregonstate.edu/confluence/display/RP/ScholarsArchive@OSU+Policies#ScholarsArchive@OSUPolicies-CollectionPolicy> or when content is better suited for a discipline-specific repository. ScholarsArchive@OSU is not an appropriate repository for data subject to FERPA, HIPAA/HITECH <https://www.hhs.gov/hipaa/> regulation of sensitive clinical or medical data, or other personally-identifiable information (PII) with disclosure risk. It is the contributor's responsibility to redact or anonymize content containing personally identifiable information prior to submission to ScholarsArchive@OSU. Datasets contributed to ScholarsArchive@OSU are monitored by a Data Curator for conformance to these obligations. Unrestricted, sensitive and confidential data are defined by OSU's Office of Information Security in the University Data Management, Classification, and Incident Response Policy <https://policy.oregonstate.edu/UPSM/08-015_university_data_management_policy>. Of these, only unrestricted data - "data intended for appropriate general use within the university" - may be eligible for deposit to ScholarsArchive@OSU. Confidential and sensitive data are not accepted in ScholarsArchive@OSU in any form, and

unrestricted data pertaining to human subjects are accepted for deposit in ScholarsArchive@OSU only following a review process. Unrestricted, sensitive, and confidential data correspond to levels 1, 2, and 3, respectively, as defined by OSU's Human Research Protection Program and Institutional Review Board <https://research.oregonstate.edu/irb/policies-and-guidance-investigators/guidance/data-security>. Researchers interested in depositing unrestricted data from human subjects must communicate the presence of human subject data during the submission process. They will work with a ScholarsArchive@OSU reviewer and, when applicable, the Office of Research Integrity, to ensure that the dataset complies with all the regulations, laws, and ethical standards applicable to the dataset.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Reviewer 3

Comments:

V. 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:

The ScholarsArchive@OSU repository is managed by staff within one of the five core departments within OSULP: the Emerging Technology and Services department. There are a total of 3 FTE with responsibilities for the management and maintenance of the repository, although additional FTE are assigned based on project and need. FTE includes the following relevant positions: Emerging Services and Technologies Department Head, Digital Repository Librarian, Scholarly Communication Librarian, Digital Services Librarian, Metadata Librarian, Metadata Technician, Operating Systems Network Analyst, a team of four programmers and two student developers. Four library faculty with data management and data curation responsibilities (referred as Data Curators), including the Data Management Specialist, a Social Sciences Data Librarian, a Coordinator for Research Services, and the Engineering Librarian, report to the more public-facing Teaching and Engagement department (soon to be renamed Research and Learning), demonstrating the importance to the library of repository-related faculty and student outreach. Several of the above positions are members of the ScholarsArchive@OSU User Group that meets each academic term to: 1. Review potential changes to SA@OSU related workflows, policies, and procedures; 2. Review and suggest changes to the repository; 3. Discuss and agree to workflows for processing new collections of materials and determine whether materials are appropriate for the repository or not.

-The Digital Repository Librarian manages the two OSU Libraries and Press repositories: ScholarsArchive@OSU and OregonDigital. The position serves as a project manager for repository development.

-The Metadata Librarian is responsible for metadata associated with the two repositories, creates and manages application profiles, and ensures metadata quality and enhancement.

- The Metadata Technician creates, reviews, and edits metadata for ScholarsArchive@OSU repository objects and ensures the accuracy, viability, and collection appropriateness of deposited content.

-The Digital Services Librarian assesses repository effectiveness in meeting operational library needs and develops tools to enhance repository workflows and content preservation.

-The Operating Systems Network Analyst oversees and implements all repository-related system hardware and infrastructure.

-The Emerging Technologies and Services programming team is responsible for writing code in support of the repository.

- The Data Management Specialist works with faculty, academic units, and students to develop and sustain services in support of curation of data produced at OSU; consults with researchers about their data, its management, reuse, and accessibility; assists researchers with creating data documentation and metadata aimed at preparing data for sharing in digital repositories; assists faculty in writing data management plans; assists faculty in meeting funder requirements for providing public access to research results, including data; is informed of, and keeps the library and university faculty up-to-date on national and international trends, legislation, and pending changes in public access requirements for data; collaborates with Emerging Technologies and Services department faculty and staff in developing funding agency data deposit services and workflows; participates in and leads campus initiatives, committees, and task forces relating to data management and storage; tracks international developments in data curation and participates in library, campus, regional, and national discussions regarding cooperative data curation activities and services.

-The Scholarly Communication Librarian serves as the library's point person and conducts outreach to OSU faculty, research assistants, staff, and students for the development and preservation of digital collections.

- The Data Curators (four librarians, including the Data Management Specialist) assist researchers with creating data documentation and metadata aimed at preparing data for sharing in ScholarsArchive@OSU, and review and curate

datasets submitted to the repository.

VI. 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

Response:

Librarians at OSU are tenure-track faculty and are actively involved in research, teaching and service activities relating to open access, data management and stewardship, and digital preservation <https://ir.library.oregonstate.edu/collections/p26771391>. For example, see the article "Peer Review of Research Data Submissions to ScholarsArchive@OSU: How can we improve the curation of research datasets to enhance reusability?" <https://doi.org/10.7191/jeslib.2019.1166> in which two OSULP faculty ask the Designated Community, researchers

wanting to reuse ScholarsArchive@OSU datasets, to review these datasets to evaluate their reusability and inform curation practices.

ScholarsArchive@OSU is represented on the University Digital Research Infrastructure Committee, responsible for accomplishing the research strategy of OSU's IT Strategic Plan and advising on the strategic direction of the university's digital research infrastructure. The Committee includes data science and disciplinary research data experts as well as information technology personnel and serves as a resource for answering questions pertaining to changes in data processing and science infrastructure. OSULP has institutional memberships in the Scholarly Publishing and Academic Resources Coalition, Coalition for Open Access Policy Institutions, Coalition for Networked Information, Archives West, Greater Western Library Alliance, Digital Library Federation, DataCite, ORCID, OCLC, the Center for Research Libraries, Orbis Cascade Alliance, and MetaArchive. The library is a partner in the Samvera repository open source community <https://samvera.org/samvera-partners/>. The Scholarly Communication Librarian was MetaArchive Steering Team Chair-Elect (2019-2020) and OSU librarians participate in monthly MetaArchive community meetings to discuss repository-related digital preservation best practices, workflows, tools, and strategies.

The amount of staffing at OSULP in support of preserving and making accessible the University's research changes according to need. ScholarsArchive@OSU has historically accepted 20-30 datasets per year. In addition to a full-time Data Management Specialist, a Social Science Data Librarian is responsible for curating social science datasets, an Engineering Librarian is responsible for curating datasets submitted by the College of Engineering, and a Coordinator of Researcher Services now shares dataset review duties with the Data Management Specialist. In-house oversight, review, and feedback for the repository is provided by the advisory ScholarsArchive@OSU User Group, as described in R5. The OSU Libraries' Digital Preservation Interest Group, whose cross-departmental membership comprises faculty and staff who are involved in digital preservation work in the library, contributes guidance on digital preservation policies and strategies for ScholarsArchive@OSU.

Reviewer Entry
Reviewer 1
Comments:
Reviewer 2
Comments:
Reviewer 3
Comments:

DIGITAL OBJECT MANAGEMENT

VII. Data integrity and authenticity

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

Compliance Level:

3 – The repository is in the implementation phase

Response:

The ScholarsArchive@OSU repository verifies the integrity of files during ingest using checksum tools. The FITS tool is run on files producing an 'original' MD5 checksum, and Fedora later computes its own SHA1 checksum. Other characterization metadata (file format, mime type, etc.) are provided by FITS. PREMIS metadata is not currently produced or stored with digital objects, although plans are in place to test the Fedora feature that generates and stores PREMIS metadata for ScholarsArchive@OSU file changes. With active preservation in mind, the repository uses Fedora's fixity service to record and periodically evaluate file integrity for every file uploaded.

Provenance metadata is automatically captured as part of the deposit process and each time metadata is revised. This provenance metadata includes the name and email address of the depositor, the name and email address of the person who approved the deposit, the time these actions were taken (e.g. 2013-08-06T17:01:01Z (GMT)), the number of bitstreams added, the number of bytes and the checksum for each, the file names, and the date that the item was made available in the repository. Date-modified metadata is stored with each file version in Fedora, but metadata is not versioned at this time.

Version control is available to users of ScholarsArchive@OSU in order to update or roll-back file versions. Creating a new version of a file results in re-characterization of the file, generation of a new thumbnail, and replacement of the downloadable file with the new version. Repository administrators have the ability to roll-back any versioning decisions in the system, as all old versions of the file are maintained within the system.

Depositors and repository technicians who are authorized to take actions on deposits are authenticated through the OSU ONID central identification system, an enterprise single sign-on (SSO) protocol. Most content added to the repository is curated and reviewed by a Metadata Technician prior to repository publication. This includes review of the completeness of the data, appropriateness of the deposit file types, the inclusion of the content in appropriate repository collections, review of the metadata, versions, licensing, and the relation of files to other materials within and external to the repository

as necessary.

For datasets, Data Curators manually review all the information that is added to dataset metadata records, check the license, ensure that the abstract (required for datasets) is understandable, and review ReadMe file documentation for accuracy and understandability. The ReadMe template <https://raw.githubusercontent.com/osulp/Data-Management-Templates-Project/master/Documentation_Template/Documentation_Template.txt> offered to researchers to facilitate documentation includes, among others, questions about provenance and versions of the dataset. In addition, the Data Curators manually review data file formats to determine if they are in the recommended formats, and where not, provide consultation to the researcher about file format preferences <https://guides.library.oregonstate.edu/c.php?g=708682&p=8342195>. As much as possible depending on the contents of the dataset, depositors are asked to provide data in non-proprietary, actionable, platform-independent formats. The Data Curators rely on guidance from the Data Curation Network primers <https://datacurationnetwork.org/outputs/data-curation-primers/> and the disciplinary expertise of the submitter to evaluate the dataset documentation and formats. For datasets with human subject data, Data Curators also confirm that there are no direct identifiers and that the Institutional Review Board approved the sharing of the dataset. File validation is a manual process at this moment, that consists of opening each file, confirming that the file can be opened, and that it contains the type of content expected by the extension. We are currently exploring tools such as DROID for automated validation.

Reviewer Entry
Reviewer 1
Comments:
Reviewer 2
Comments:
Reviewer 3
Comments:

VIII. 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
Accepted as amended.

Reviewer 2

Comments:

Reviewer 3

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

Response:

The ScholarsArchive@OSU Collection Policy <https://wiki.library.oregonstate.edu/confluence/display/RP/ScholarsArchive @OSU+Policies#ScholarsArchive@OSUPolicies-CollectionPolicy> specifies the types of content that are accepted into the repository. Examples of possible content to be deposited include, but are not limited to: faculty research articles; dissertations and theses; reports, whitepapers, and technical documents; award-winning student papers or presentations; conference papers; journals and books published by OSU; learning objects; faculty supported student works; research datasets; creative works; OSU sponsored or affiliated conference proceedings; publications and reports from affiliated organizations.

A prescribed set of descriptive metadata fields is defined for each type of work to be deposited in ScholarsArchive@OSU, and these definitions are made accessible to depositing scholars within the repository's deposit forms. A title, creator, resource type, and rights statement metadata are required to be entered by the depositor for all resource types, with entries for the latter two fields selected by the depositor from a controlled vocabulary. Deposits do not reach the deposit review stage without information in these fields. A reviewer manually conducts quality control checks of metadata and files after deposit. Repository staff reserve the right to edit user-submitted metadata in order to correct or normalize values, to make updates in keeping with changing metadata standards, or to enhance descriptions for greater discoverability. Technical and administrative metadata -- including depositor name, date uploaded, date modified, file format, MIME type, and additional file characterization information that varies with file format -- are extracted and generated automatically by the repository software at the time of deposit and are stored in the repository.

In cases where the reviewer requires additional metadata pertaining to the object, the depositor is notified that additional metadata is required before the object is approved for publication in the repository. In cases where the data itself is inadequate to be launched, the data is rejected by the reviewer and returned to the data provider for rectification and redeposit.

To ensure that deposited datasets are usable by others, ScholarsArchive@OSU requires the creation of a documentation file(s) for all datasets, submitted in a non-proprietary format (e.g. .txt, .pdf) for recording and sharing data-level metadata. No formal metadata standard is in place beyond the generic ReadMe template <https://raw.githubusercontent.com/osulp/ Data-Management-Templates-Project/master/Documentation_Template/Documentation_Template.txt>, which is available to depositors for modification and use. Researchers are encouraged to adopt a research metadata standard such as Ecological Metadata Language (EML) for documenting their data, but the use of such research metadata standards is not required. This documentation file is submitted and stored alongside the data files and is distinct from the descriptive

metadata describing the dataset as a whole.

Files must be presented, wherever possible, in formats that are non-proprietary (e.g. CSV should be used instead of XSLX). Files must be actionable and usable in an analysis application (e.g. tabular data should never be shared in pdf or image formats), platform-independent, and accessible in all operating systems. Preferred file formats for submission to the repository are outlined in the Preferred File Formats document within the ScholarsArchive@OSU help page <https://guides.library.oregonstate.edu/c.php?g=708682&p=8342195>. Based on the initial round of Core Trust Seal reviewer comments, ScholarsArchive@OSU now provides information to depositors about required file formats for textual documents such as theses and dissertations, research articles, presentations, technical reports, conference proceedings, posters, etc. The PDF file format is required for this content. PDF/A-1 -- ISO 19005-1 with fonts embedded (.pdf) is preferred. For other content types such as quantitative and statistical data, spreadsheets, databases, graphics, audio, and video (among others) we provide a table of appropriate and recommended file formats for preserving and sharing digital files in ScholarsArchive@OSU.

ScholarsArchive@OSU will accept file formats that do not comply with the file format recommendations only after consultation with the depositor, when recommended formats are not available for the file type, or when formatting of the recommended file format makes the data file substantially less usable. In these cases, non-compliant formats may be submitted alongside recommended formats to facilitate near-term reuse of data.

ScholarsArchive@OSU is committed to providing bit-level preservation of all content contained in the repository regardless of the file format. OSU Libraries and Press guarantees that objects submitted to ScholarsArchive@OSU using recommended file formats will be retrievable in the future. This covers common file formats including PDF (.pdf), XML (.xml), Text (.txt), HTML (.htm, .html), JPEG (.jpg, .jpeg), GIF (.gif), and TIFF (.tif, .tiff), and other formats designated as "Highest Confidence" by the ScholarsArchive@OSU Preferred File Formats document <https://guides.library.oregonstate.edu/c.php?g=708682&p=8342195>. However, users are informed in the ScholarsArchive Policies <https://wiki.library.oregonstate.edu/confluence/x/bivWAw> that the proprietary or executable nature of some file formats may make it impossible to guarantee persistent access to those file formats. As the need arises, ScholarsArchive@OSU staff and the OSU Libraries' Digital Preservation Interest Group will continue to determine best strategies for ongoing retrieval and understandability.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Reviewer 3

Comments:

IX. 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

Response:

ScholarsArchive@OSU repository policies and procedures are managed by the Digital Repository Librarian and the Metadata Technician. Additionally, data curation procedures are managed by the Data Management Specialist. Procedure documentation is located on an internal shared drive accessible to repository staff. Policies are posted to a public space for repository policies <https://wiki.library.oregonstate.edu/confluence/x/bivWAw>. Changes or additions to policies are approved by the ScholarsArchive@OSU User Group, and, in some cases, vetted by OSU Libraries Administration, Management and Planning group (LAMP) before becoming codified.

All data is stored on HDD/SSD locally, and AWS S3 is used for off-premises storage. These storage systems monitor individual file objects via checksum and auto-repair. Fedora's fixity service is used to monitor the consistency of repository objects, and AWS runs fixity checks to verify individual object consistency within backups. As noted in the ScholarsArchive@OSU Preservation Policy, fixity reports are reviewed at least annually and relevant files are restored from backups as needed. No optical discs or magnetic tapes or other media that breaks down physically over time are used for storing Scholars Archive content, locally or off-premises.

ScholarsArchive@OSU relies on Amazon Web Services (AWS) S3 <https://aws.amazon.com/s3/> for backups. AWS complies with a number of regulatory and professional IT standards and certification programs <https://aws.amazon.com/compliance/>, including CSA, FERPA, FISMA, HIPAA, ISO 9001, 2701, 2017, SOC 1, 2, 3, and others. The service level agreements defining the terms of the contractual relationship between OSU and Amazon are available at <https://aws.amazon.com/agreement/>, <https://aws.amazon.com/s3/sla/>, and <https://aws.amazon.com/ec2/sla/>.

Internal documentation that outlines archival storage for all ScholarsArchive@OSU content is maintained. Backups of the filesystem, individual repository objects, and relational database are maintained on-device (campus server), off-device in the cloud (AWS) and off-site in another timezone (AWS). Backups are run nightly and backups are compared with the originals nightly. Additionally, the source storage has point-in-time snapshots hourly, daily, and weekly for all components. Repository level exports provide a mechanism for simple, consistent restores, including regeneration of search indexes. Declarative infrastructure allows us to quickly restore services that present the data.

MetaArchive <http://metaarchive.org>, as described in R10, is used only for the selective replication of repository content that is deemed to be most at risk and of the greatest historical importance to the university: the university's corpus of Electronic Theses and Dissertations and Extension and Experiment Station Communication publications. MetaArchive uses LOCKSS <https://www.lockss.org/community/networks/> to ensure consistency across archival copies. Data is stored in five distributed digital preservation nodes. LOCKSS regularly and iteratively checks with each of the copies to make sure that the contents are identical over time. More information about MetaArchive's distributed digital preservation system can be found at <https://metaarchive.org/how-ddp-works/>.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Reviewer 3

Comments:

X. 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
Accepted as amended.

Reviewer 2

Comments:

Reviewer 3

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

Response:

The OSU Libraries' Digital Preservation Policy
<https://wiki.library.oregonstate.edu/confluence/display/RP/OSU+Libraries+Digital+Preservation+Policy> makes explicit the ScholarsArchive@OSU repository's commitment to preserving its digital resources through a comprehensive digital preservation program.

As described in the ScholarsArchive@OSU Preservation Policy <https://wiki.library.oregonstate.edu/confluence/display/RP/ScholarsArchive@OSU+Policies#ScholarsArchive@OSUPolicies-PreservationPolicy>, the repository provides ongoing support to ensure long-term viability for as many file formats as possible. OSULP is committed to maintaining and preserving access to the content and metadata in ScholarsArchive@OSU. ScholarsArchive@OSU is committed to providing bit-level preservation of all content. After the comments from the last round of reviews for the CTS certification, we have revised the ScholarsArchive@OSU Preservation Policy so that it includes an explicit commitment to guarantee the retrievability of recommended file formats, and we have expanded the recommended file formats to include other repository content types such as still images, video, audio, and textual documents. This covers common file formats including PDF (.pdf), XML (.xml), TXT (.txt), HTML (.htm, .html), JPEG (.jpg, .jpeg), GIF (.gif), and TIFF (.tif, .tiff), and other formats designated as "Highest Confidence" by the ScholarsArchive@OSU Preferred File Formats document <https://guides.library.oregonstate.edu/Scholars-Archive/PreferredFileFormats>. "Highest confidence" formats maximize preservation by prioritizing formats that are non-proprietary (e.g. CSV should be used instead of XLSX), actionable and usable in an analysis application (e.g. tabular data should never be shared in PDF or image formats), platform-independent, and accessible in all operating systems. In the case of deprecation or obsolescence of any recommended file format(s), OSU digital repository staff will engage in format migration or other active preservation strategies as appropriate to provide continuing access to and usability of the objects.

ScholarsArchive@OSU is format-agnostic and will accept file formats that are not in the "Highest Confidence" list where other formats are not available or where formatting makes the data file substantially less usable. Non-compliant formats may be submitted alongside compliant formats to facilitate near-term reuse of data. However, OSULP offers consultation and guidance on ways to create digital content in a manner that is most amenable to the highest level of future preservation service <https://guides.library.oregonstate.edu/c.php?g=708682&p=8342195>. Only bit-level preservation is ensured for non-preferred formats.

ScholarsArchive@OSU maintains a complete change history of managed content by automatically logging versions and changes associated with items in the repository.

The most recent revision (May 2022) of the ScholarsArchive@OSU Preservation Policy <https://wiki.library.oregonstate.edu/confluence/display/RP/ScholarsArchive@OSU+Policies#ScholarsArchive@OSUPolicies-PreservationPolicy> also describes annual assessments of our preservation practices that include (1) a review of the preferred file formats

document to ensure that recommendations remain up to date with current national and international standards and best practices. This review will be informed by resources such as the Library of Congress's Recommended Formats Statement and the Open Preservation Foundation's International Comparison of Recommended File Formats. (2) An annual inventory of file formats stored in the repository. This step includes a plan to address format migration(s) for file formats deemed to be in jeopardy. (3) A review of the ScholarsArchive@OSU Preservation Policy that will take into account the evolving needs of the ScholarsArchive@OSU depositors and its Designated Community of users. (4) A review of fixity reports and restoration of relevant files from backups as needed. Any significant changes to policy or practice resulting from this annual assessment are approved by the ScholarsArchive@OSU User Group, as well as the OSU Libraries' Digital Preservation Interest Group as appropriate.

ScholarsArchive@OSU relies on Amazon Web Services (AWS) S3 for content replication and backups. In addition, MetaArchive is also used as a backup for repository content that is deemed to be most at risk and of the greatest historical importance to the university. The MetaArchive approach <http://www.metaarchive.org/> uses the LOCKSS software <http://www.lockss.org>. ScholarsArchive@OSU relies on MetaArchive for replicating the two largest collections of content in the repository--the university's corpus of Electronic Theses and Dissertations and Extension and Experiment Station Communication publications--at five different geographically dispersed servers around the globe. The five preservation nodes regularly and iteratively check in with each other to make sure that all five copies of the content remain identical over time.

OSU Libraries and Press became a sustaining member of the MetaArchive Cooperative in 2010. As a sustaining member of MetaArchive, OSU Libraries is responsible for hosting and maintaining a MetaArchive-LOCKSS cache for the member community. The membership agreement defining the terms of the relationship between OSU and MetaArchive is available upon request. In the coming year, ScholarsArchive@OSU is expected to take advantage of a free storage allotment granted to members that are hosting a LOCKSS cache to also begin ingesting ScholarsArchive@OSU datasets to MetaArchive.

MetaArchive is a dark archive; no access is provided to the archive's contents except when needed to restore a content contributor's collections. The MetaArchive Cooperative carried out a Trusted Repositories Audit & Certification (TRAC) self-assessment in 2009. A checklist report details its conformance and the evidence cited for each criterion addressed: <https://metaarchive.org/wp-content/uploads/2017/03/ma_trac_audit.pdf>.

All depositors to ScholarsArchive@OSU must agree to a non-exclusive distribution license <https://ir.library.oregonstate.edu/agreement> that grants Oregon State University the non-exclusive right to reproduce, translate, make copies for the purpose of security, backup, and preservation, and distribute a submission (including the abstract) worldwide in print and electronic format and in any medium. Depositors must agree that OSU may, without changing the content, translate the submission to any medium or format. Depositors represent that the submission is their original work and that they have the right to grant the rights contained in the license. The depositor must also represent that the submission does not, to the best of their knowledge, infringe upon anyone's copyright.
The public ScholarsArchive@OSU Withdrawal, Replacement, and Access Restriction Policy includes information on removal and withdrawal of content subject to Digital Millennium Copyright Act (DMCA) takedown notices <https://wiki.libra

ry.oregonstate.edu/confluence/display/RP/ScholarsArchive@OSU+Policies#ScholarsArchive@OSUPolicies-Withdrawal, Replacement,andAccessRestrictionPolicy>. Content deposited in ScholarsArchive@OSU will not be removed from the repository. Under certain circumstances items may be withdrawn, replaced, or restricted from view, but the content will still remain in the repository to meet our commitments to persistence and transparency. DMCA takedowns represent an exception to this removal policy. For DMCA takedown requests, the item in question will be made private but remain in the repository until final resolution of the takedown request. Upon resolution, if the decision is made to affect the takedown, the item will be removed entirely from the repository.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Reviewer 3

Comments:

XI. 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
Accepted as amended.

Reviewer 2

Comments:

Reviewer 3

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

Response:

User guidance for entering descriptive metadata is provided to depositors in the ScholarsArchive@OSU User Guide <https://guides.library.oregonstate.edu/Scholars-Archive> within designated sections for different types of deposits; for example, the guidance for Research Datasets: <https://guides.library.oregonstate.edu/Scholars-Archive/Datasets#Describe>. This resource serves as the primary user-facing instructions for entering the descriptive metadata that will be indexed in the ScholarsArchive@OSU repository; depositors are encouraged to consult with Data Curators if they are unsure how to use any metadata field. The complete Metadata Application Profile for the repository, used primarily for internal documentation and configuration, is publicly available on Google Drive <https://docs.google.com/spreadsheets/d/1koKjV7bjn7v4r5a3gsowEimljHiAwbwuOgjHe7FEtuw/edit?usp=sharing>. Each resource type (e.g. Datasets, Articles, Theses and Dissertations, Technical Reports, etc.) defined in the ScholarsArchive@OSU repository has a relevant set of metadata properties. The Metadata Application Profile defines the full set of properties for all resource types in the MAP tab, and the specific metadata properties for a given resource type, such as Datasets, are found in the Metadata Ordering tab; note that these terms correspond to the properties' Form Labels (in column I of the MAP tab) rather than property names. Mandatory properties -- title, creator, resource type, and rights statement -- are shown in bold. A ReadMe tab is included in the Metadata Application Profile, explaining the purpose of the columns in the MAP tab. Several metadata elements are controlled and others recommend the use of controlled terms. Various mechanisms are provided to assist depositors with selecting from controlled vocabularies. These include both system interventions such as drop-down menus and type-ahead tools, as well as staff assistance such as consultations on the use of Library of Congress Subject Headings. Links to related works within the repository and external to it are provided as appropriate and at the discretion of the depositor.

ScholarsArchive@OSU requires the creation of a documentation file(s) for all datasets, and provides a generic ReadMe template for this documentation file <https://raw.githubusercontent.com/osulp/Data-Management-Templates-Project/master/Documentation_Template/Documentation_Template.txt>. This documentation allows depositors to record information about the submitted data beyond the scope of the repository's descriptive metadata properties. The documentation file can be used by researchers to enhance understanding and assessment of datasets in the repository.

After content is deposited to the repository, it is reviewed for quality by the Metadata Technician or a Data Curator depending on the content type. The quality check is carried out to ensure that the data, metadata, and documentation meet the requirements specified in the Research Data Curation and Collection Policies <https://wiki.library.oregonstate.edu/confluence/x/bivWAw>. If the data package does not meet the requirements, the depositor will be contacted by email to ask for improvements.

The quality check includes controlling the following aspects pertaining to readability and accessibility:
-Are the files delivered in an accepted file format?
-Are the files readable or saved in a portable format?
-Do the files fall within the maximum data limit?
-Is there adequate documentation about the data and supplementary data? (Data Report template is provided to the depositors)
-In case of several files, is the folder structure clear to you, and are all files included?

-Are the data files complete?

-Is the data free of any privacy sensitive information?

XII. 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

Response:

Data are ingested to the repository by eligible users. Flexible access controls that allow depositors to restrict access to the material are offered (see R2 response for details). Data are ingested to the repository using the Hyrax API via a Rails application that interfaces directly with the Hyrax software. The repository software automatically routes deposited works to administrators and stakeholders for review, with specific reviewer roles designated within the system according to the type of work being deposited.

OSU researchers are not required to submit their datasets to ScholarsArchive@OSU. ScholarsArchive@OSU is available for any OSU researcher, if they need it. The Data Curators consult with researchers when appropriate, and discuss repository options and where to preserve and share their data. Alternative disciplinary repositories are encouraged, as are repositories recommended by funders.

Each deposit undergoes file characterization and queuing for backup. All metadata is manually reviewed by a Metadata Technician or a Data Curator and files are randomly checked to ensure they are actionable prior to publication. An email that includes the persistent URL is delivered to the depositor automatically upon repository publication.

The workflow for textual documents is maintained internally. The ScholarsArchive@OSU User Group, an advisory group that includes the Data Management Specialist, Metadata Librarian, Digital Repository Librarian, Emerging Technologies and Services Department Head, Scholarly Communication Librarian, Analyst Programmer, and a Metadata Technician, meets each academic term to review potential changes to workflows, policies, and procedures. Other staff are invited based on need.

The workflow for datasets followed by the Data Curators is likewise outlined in an internal document. Updates on the workflow are discussed and agreed upon by curators, and when necessary the ScholarsArchive@OSU User Group is consulted.

The dataset workflow documentation describes the following process. Each dataset submission is initiated by the submitter, and requires filling out all the required fields in ScholarsArchive@OSU, attachment of at least one file, and agreeing to the repository license. If files are larger than 500MB, the upload must be done by the Data Curator since the ScholarsArchive@OSU interface does not allow the upload of files that large. The submitter receives an automatic e-mail from ScholarsArchive@OSU when the submission has been completed successfully. The Data Curator also receives an automatic notification, and sends a response within 24 hours of submission via ScholarsArchive@OSU acknowledging that the dataset has been received, informing the submitter about the review process, providing the submitter with a temporary DOI for the dataset, and asking if there are deadlines to approve the submission. The curator then reviews the dataset following the criteria in the workflow document, and sends a message to the submitter with the list of changes necessary to approve the dataset, indicating clearly those that are mandatory and those that are only recommended. The Data Curator and the submitter exchange as many messages (via e-mail or via the ScholarsArchive@OSU messaging system) as needed until the dataset is ready for approval. The Data Curator submits the relevant metadata to Datacite to generate a permanent DOI, adds the DOI to the metadata record, and approves the dataset in ScholarsArchive@OSU. The submitter receives an automatic e-mail.

Separate but similar internal documentation exists describing the workflows for other content types.

During the review process a Data Curator checks file formats and completeness of documentation and record metadata. A Data Curator also checks that the data is appropriate for ScholarsArchive@OSU and that it can be shared.

ScholarsArchive@OSU only accepts unrestricted data (as defined by OSU's University Data Management, Classification, and Incident Response Policy <https://uit.oregonstate.edu/ois/data-management-and-classification-overview>. When the Data Curator detects that the dataset includes human subjects data, they follow a human subjects protocol to ensure that the project was reviewed by IRB, that the data is unrestricted, and that it can be shared according to the approved protocol and the consent forms. The human subjects protocol for ScholarsArchive@OSU was created in collaboration with Oregon State University's Institutional Review Board.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Reviewer 3

Comments:

XIII. 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
Accepted as amended.

Reviewer 2

Comments:

Reviewer 3

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

Response:

ScholarsArchive@OSU provides for direct search and retrieval of repository objects from the home page <https://ir.library.oregonstate.edu>. Advanced search options include the ability to search within specific fields and across

multiple fields with a single search. Search help is provided for accessing specific types of resources: <https://guides.library.oregonstate.edu/Scholars-Archive/Search>. In addition, all content housed within the ScholarsArchive@OSU repository is discoverable in Google Scholar, Google, and other internet search engines.

ScholarsArchive@OSU persistent URLs assigned to every item in the repository are shown prominently at the top of item landing pages. DOIs are allocated and registered through the DataCite DOI registration agency for all datasets housed in the repository. The DOI is displayed in the DOI metadata field on dataset landing pages. At present, the DataCite submissions are created manually using an internally documented mapping of ScholarsArchive@OSU metadata properties to the DataCite schema. Given the steadily increasing number of dataset submissions to the repository, OSULP is investigating automated crosswalking and metadata submission to DataCite. ScholarsArchive@OSU datasets are discoverable through DataCite's search service <https://search.datacite.org> and DataCite Commons <https://commons.datacite.org>, and are exposed to aggregator services through DataCite's APIs. For example, OSF SHARE <https://share.osf.io/> indexes ScholarsArchive@OSU datasets for which DOIs have been registered through DataCite.

Repository content other than datasets, such as research articles and technical reports, may have DOIs registered with the Crossref DOI registration agency. Crossref submissions are also created manually, typically at the request of the author; again, the DOI is displayed in the DOI metadata field on the repository record for the work. Documents registered with Crossref are discoverable through Crossref's search service <https://www.crossref.org>.

Descriptive metadata stored in the ScholarsArchive@OSU repository is transformed to OAI-compliant Dublin Core documents according to the Dublin Core/Qualified Dublin Core mappings shown in the Metadata Application Profile <https://docs.google.com/spreadsheets/d/1koKjV7bjn7v4r5a3gsowEimIjHiAwbwuOgjHe7FEtuw/edit#gid=0>. Repository metadata is harvested using OAI-PMH for search and retrieval in the Oregon Explorer natural resources digital library <http://oregonexplorer.info/>.

OSULP recommends a standard citation format for datasets housed in the repository and provides data depositors with instruction in creating a citation <https://guides.library.oregonstate.edu/research-data-services/data-management-data-citation>. A recommended citation is included in the ReadMe documentation file for datasets (it is a field in the ReadMe template). Landing pages in ScholarsArchive@OSU also feature pre-formatted citations conforming to the 2014 FORCE11 Joint Declaration on Data Citation Principles <https://www.force11.org/group/joint-declaration-data-citation-principles-final>. For example:

Clark, D., & Wilson, M. V. (2019). The Willamette Valley (Oregon) Prairie Plant Trait Dataset (Version 1) [Data set]. Oregon State University. https://doi.org/10.7267/6q182r738

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Reviewer 3

Comments:

XIV. 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
Accepted as amended.

Reviewer 2

Comments:

Reviewer 3

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

Response:

The ScholarsArchive@OSU Metadata Application Profile describes required and suggested metadata labels, predicates, field vocabularies and other properties for creating bibliographic metadata for all repository work types <https://docs.google.com/spreadsheets/d/1koKjV7bjn7v4r5a3gsowEimljHiAwbwuOgjHe7FEtuw/edit#gid=0>. ScholarsArchive@OSU employs metadata elements from the Dublin Core Metadata Element Set and other established vocabularies expressed as RDF predicates. Mandatory metadata elements for all work types are title, creator(s), rights statement, and resource type. The date uploaded, depositor, and date modified elements are captured automatically, as is file characterization information. Other optional elements--funder, keywords, abstract, methodology, usage notes, related datasets/publications, and location (point, bounding box, or place name)--may also be supplied for datasets, and their use is strongly encouraged. Use of linked data for controlled fields clearly defines the meaning of those fields, contributing to machine and human reuse and understandability. Repository work types are defined and persisted for the purpose of declaring metadata fields and requirements for type of work in the repository. Work type examples include: article, dataset, technical report, open educational resource, etc.

The repository exposes structured metadata for reuse via the Open Archives Initiative Protocol for Metadata Harvesting. The Metadata Application Profile specifies the mappings of ScholarsArchive@OSU properties to the Dublin Core and Qualified Dublin Core schemas to provide OAI-compliant metadata for harvesting by external services and organizations. Metadata for datasets is also crosswalked to the DataCite 4.0 schema <http://schema.datacite.org/> for DOI registration according to separate internal documentation.

In addition to the descriptive metadata indexed in the repository, described above, ScholarsArchive@OSU requires the creation of a documentation file(s), submitted in a non-proprietary format (e.g. TXT or PDF) for recording and sharing data-level metadata in order to make deposited datasets usable by others. This data-level documentation is intended to be more granular and flexible than the bibliographic metadata record. A generic ReadMe template <https://raw.githubuser content.com/osulp/Data-Management-Templates-Project/master/Documentation_Template/Documentation_Template.txt> is available to depositors for modification and use. Mandatory information required in the documentation file is: Title and abstract for the dataset (duplicated from the repository record); Names of creators (consistent with the repository record, with additional information); Name and contact information of a contact person; License or explanation of restrictions placed on the data (consistent with the repository record); Methods of data collection; A file overview with a description of all the files included in the dataset; Description of the data elements presented in the data file/s. For tabular data this means a data dictionary that will include a description of all variables. In addition, OSULP strongly suggests that contributors provide additional data-level metadata for their dataset that includes: Data origin (experimental, observational, raw or derived, physical collections, models, images, etc.); Data type (integer, Boolean, character, floating point, etc. Instrument(s) used); Data acquisition details (sensor deployment methods, experimental design, sensor calibration methods, etc.); File type (CSV, mat, xlsx, tiff, HDF, NetCDF, etc.); Data processing methods, software used; Data processing scripts or codes; Dataset parameter list, including Variable names; Description of each variable; and Units <https://guides.library.oregonstate.edu/research-data-services/data-management-metadata>.

A Data Curator reviews all the information that is added to dataset metadata records, checks the license, ensures that the abstract (required for datasets) is understandable, reviews data files to make sure they are in the right formats, and reviews ReadMe file documentation. Inconsistencies between metadata in the ReadMe documentation versus the bibliographic metadata provided for the deposit are raised and resolved in communication with each depositor.

As much as possible depending on the contents of the dataset, depositors are asked to provide data in non-proprietary, actionable, platform-independent formats. As described in R10, ScholarsArchive@OSU guarantees the ongoing usability of repository objects submitted using recommended file formats, and will migrate those recommended formats as needed to meet this commitment.

Reviewer Entry
Reviewer 1
Comments:
Reviewer 2
Comments:

Reviewer 3

Comments:

TECHNOLOGY

XV. 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:
Accept

Reviewer 2

Comments:

Reviewer 3

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

Response:

ScholarsArchive@OSU uses the Samvera Hyrax digital asset management solution. It is an open source, community-supported platform <https://samvera.org/samvera-flexible-extensible/the-samvera-community/> used primarily by libraries, archives, and museums. Samvera Hyrax adheres to international library metadata and other standards and to recognized "best practice" for design and development. The primary components of Samvera Hyrax utilized by OSULP are a Fedora Commons content backend, Solr for indexing, SQL databases, and Blacklight for search and retrieval.

The Emerging Technologies and Services department at OSULP runs ScholarsArchive@OSU and is responsible for maintaining the repository's technology stack. The Director of the Emerging Technologies and Services Department, a member of the ScholarsArchive@OSU User Group with the Data Management Specialist, the Digital Repository Librarian, and others, directs and oversees repository technology maintenance. The Emerging Technologies and Services

department has an infrastructure development plan that includes running the repository in our on-premises Kubernetes cluster with continual replication to AWS S3 for long term backups. MetaArchive is used for digital preservation of selected content. A software inventory and system documentation are maintained by the department for internal use. OSU's Internet connectivity is provided through multiple peering relationships including the 100Gbps Internet2 network. Internet bandwidth has proven sufficient to meet user needs and requirements for round-the-clock repository connectivity and usage. All technologies in use are reviewed at least annually and as needed by an Emerging Technologies and Services department technology manager and two infrastructure analysts.

ScholarsArchive@OSU employs metadata elements from the Dublin Core Metadata Element Set and other established vocabularies expressed as RDF predicates. All ScholarsArchive@OSU metadata is serialized as RDF and stored as RDF statements. RDF as a model is governed by W3C and the RDF Working Group <https://www.w3.org/RDF/>. Fedora also adheres to the W3C Linked Data Platform specification <https://www.w3.org/TR/ldp/>. Details about the ScholarsArchive@OSU RDF graph implementation are available on request.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Reviewer 3

Comments:

XVI. Security

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

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase
Accepted as amended.

Reviewer 2

Comments:

Reviewer 3

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

Response:

The ScholarsArchive@OSU institutional repository lives on Oregon State University-administered and maintained VMware ESX and Dell PowerEdge servers. The University provides a base 64bit Linux Operating System on VMware ESX hosts along with VMware HA (High Availability) on Dell PowerEdge servers connected to 4GB fiber array disks. The University's Information Services unit takes snapshots of our VM before updates are applied or applications are installed to test. The use of virtualized servers and containers orchestrated on physical servers in a hybrid fashion provides the scalability and flexibility of virtualization, with the performance of physical servers. They allow maintenance and upgrades to the background hardware without incurring downtime, hardware high availability due to redundant backend hardware, and faster server deployment since there is no need to purchase, install or configure hardware. Virtualized servers also support the University's campus green initiative, saving power, cooling and data center space. Physical servers are housed in a locked and climate-controlled university facility. Access to computing facilities is controlled by university Information Services. Only personnel that need access are granted access, and there are established university procedures for this.

Oregon State University Information Systems <https://is.oregonstate.edu/> provides a robust, scalable, and adaptable technology ecosystem that provides continuity and reliability, ascertaining and managing risks to the university's data and infrastructure. The unit restricts system access and availability to the repository using the ONID authentication system. The backend is further restricted to two system administrators within the library. The Office of Information Security is tasked with identifying threats to the data housed in the repository, such as hackers and the malicious software they use. Employees responsible for disaster preparedness and security include the following positions in the Information Services unit:
-Assistant Provost for Information Security
-Director, Oregon Research and Teaching Security
-Three other positions reporting to the Assistant Provost

The University Data Management, Classification, and Incident Response Policy <https://leadership.oregonstate.edu/sites/leadership.oregonstate.edu/files/08-015_univ_data_mgmt_policy.pdf> identifies the various types of information at OSU, the means by which information is identified as requiring protection, the protections required, and the process that happens when a failure of those protections occurs. Additional information on the University Data Management, Classification, and Incident Response policy is available here: <https://uit.oregonstate.edu/ois/data-management-and-classification-overview>.

ScholarsArchive@OSU relies on Amazon Web Services (AWS) for backups, using S3 <https://aws.amazon.com/s3/>. AWS complies with a number of regulatory and professional IT standards and certification programs <https://aws.amazon.com/compliance/>, including CSA, FERPA, FISMA, HIPAA, ISO 9001, 2701, 2017, SOC 1, 2, 3, and others. The service level agreements defining the terms of the contractual relationship between OSU and Amazon are available at <https://aws.amazon.com/agreement/>, <https://aws.amazon.com/s3/sla/>, and

<https://aws.amazon.com/ec2/sla/>. As noted in R0, R9, and R10, MetaArchive is used for digital preservation of repository content that is deemed to be most at risk and of the greatest historical importance to the university. MetaArchive is powered by LOCKSS software which is suitable for data preservation. MetaArchive relies on an extensive distributed digital preservation network that currently includes 15 secure, closed-access preservation nodes. Technical specifications are available: <https://metaarchive.org/wp-content/uploads/2017/03/ma_technicalspecifications.pdf>.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Reviewer 3

Comments:

APPLICANT FEEDBACK

Comments/feedback

These requirements are not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response:

Engaging in this assessment process for CTS certification has raised a number of questions about our work to ensure that our processes, staffing, and infrastructure are appropriate. There is more work that our campus and the library needs to do, especially in the areas of security and ensuring data authenticity and integrity, but we recognize that digital preservation is a process.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Reviewer 3

Comments: