

ScholarsArchive@OSU Repository

Core Trust Seal Self Assessment: June 2019

Michael Boock, Cara Key, Clara Llebot, Margaret Mellinger, Steve Van Tuyl

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type

Institutional repository, Publication repository, Library/Museum/Archives

Comments: ScholarsArchive@OSU <<https://ir.library.oregonstate.edu/>> is Oregon State University's (OSU) institutional repository for the research, scholarship, and publications of the university. OSU Libraries and Press (OSULP) is responsible for collecting, maintaining, preserving, and providing access to the items in ScholarsArchive@OSU. Repository holdings include all 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 missions of the university.

Brief Description of the Repository's Designated Community

All members of the OSU community--faculty, students, and staff in all disciplines--are eligible to contribute to the ScholarsArchive@OSU repository. 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.

Level of Curation Performed.

B. Basic curation – e.g., brief checking, addition of basic metadata or documentation

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 university's open access policy. These are made available without undergoing review.

A Data Management Specialist reviews 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. The Specialist provides 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 the Specialist prior to publication. A ReadMe template is available to researchers from https://guides.library.oregonstate.edu/ld.php?content_id=45294345

Outsource Partners:

OSULP relies on the MetaArchive Cooperative for content replication and preservation services. OSU became a sustaining member of the MetaArchive Cooperative in 2010. The MetaArchive approach <<http://www.metaarchive.org/>> relies upon a distributed digital preservation network infrastructure that uses the LOCKSS software <<http://www.lockss.org>>. OSULP replicates 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 seven different geographically dispersed servers around the globe.

The seven preservation nodes regularly and iteratively check in with each other to make sure that all seven copies of the content remain identical over time. OSULP is completing the testing of MetaArchive infrastructure to replicate remaining ScholarsArchive@OSU content including datasets. As a sustaining member of MetaArchive, OSULP 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.

As an original participant in the Orbis Cascade Alliance DataCite service, ScholarsArchive@OSU uses DataCite <<https://www.datacite.org/>> for the registration of DOIs for datasets. Prior to 2018, OSU used the California Digital Library's EZID service <<http://ezid.cdlib.org>> for assigning, managing, and resolving persistent DOIs for all repository datasets. OSU's DOI prefix is 10.7267.

OSULP also relies on external service providers for identifier management, storage, and VM hosting. ScholarsArchive@OSU relies on ORCID <<https://orcid.org/>>, an international non-profit membership organization for managing unique, persistent researcher identifiers, and Crossref's Fundref <<http://www.crossref.org/fundingdata/>>, for funding agency identifiers. ORCID and Fundref data are available via open public APIs (<<https://members.orcid.org/api/about-public-api>>, <<http://api.crossref.org>>) requiring no prior contractual relationships.

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 historic and contemporary intellectual work. ScholarsArchive@OSU also includes materials from outside the institution in support of 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 and access 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 70971 individual files totaling over 3.0 TB, expressed in 18 unique MIME types (as of June 13, 2019). As of March 20, 2019, ScholarsArchive@OSU hosts 72 datasets. OSULP strives to maintain the repository in an open and technologically neutral manner to take advantage of an evolving digital infrastructure and an international community of open source repository software developers.

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

First approved in 2015 and revised in 2019, the OSULP Digital Preservation Policy <<https://cdss.library.oregonstate.edu/sites/default/files/osulpdigitalpreservationpolicy.pdf>>, describes the 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 data curation activities of OSU Libraries and Press result, in part, 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.

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

Response: Content contributed to ScholarsArchive@OSU is covered by the terms of a license agreement granting to the University the non-exclusive right to store, display, reproduce, make certain derivative works, and publish work for the purpose of dissemination in conformance with the visibility level that is selected upon submission

<<https://ir.library.oregonstate.edu/agreement?locale=en>>.

The visibility of ScholarsArchive@OSU content 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 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 <<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 article 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. Licenses are captured and stored as part of the metadata for the repository record.

The ScholarsArchive@OSU terms of use state that repository content may be used in any manner not prohibited by copyright or other applicable law

<<https://osulibrary.oregonstate.edu/sa-termsofuse>>. OSU does not actively monitor usage to identify instances of noncompliance. 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 the Data Specialist for conformance to these obligations. Unrestricted and sensitive data are accepted for deposit in ScholarsArchive@OSU, following a review process. Confidential data are not accepted in ScholarsArchive@OSU in any form. Researchers interested in depositing sensitive data must communicate that the dataset is sensitive during the submission process. They will work with the 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, in order to assign the appropriate visibility level.

III. 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

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 earned a total of \$382 million in research funding in the fiscal year ending June 30. 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; and 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>). 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. Already, 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>>.

The Libraries' Digital Preservation Policy <<https://cdss.library.oregonstate.edu/sites/default/files/osulpdigitalpreservationpolicy.pdf>> makes explicit the ScholarsArchive@OSU repository's commitment to preserving its digital resources through a comprehensive digital preservation program. Details of digital preservation implementation for specific types of digital content will be outlined in implementation plans that are expected to be completed by Spring 2020.

For much of the content residing in the repository, university stakeholders are responsible for, or involved in, the creation and maintenance of collections. They are responsible for selecting content, within the bounds of the ScholarsArchive@OSU policies pertaining to collection development, metadata, workflows, and licensing
<<https://wiki.library.oregonstate.edu/confluence/x/bivWAw>>.

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

Metadata staff within the department are responsible for reviewing, assessing, and enhancing metadata associated with all items in the repository. A Data Management Specialist, working closely with the staff of the Emerging Technologies and Services department responsible for managing the repository, has 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 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 Management Specialist for conformance to these obligations. Unrestricted and sensitive data are accepted for deposit in ScholarsArchive@OSU only following a review process. Confidential data are not accepted in ScholarsArchive@OSU in any form. Researchers interested in depositing sensitive data must communicate that the dataset is sensitive during the submission process. They will work with the 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, in order to assign the appropriate visibility level.

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

The ScholarsArchive@OSU repository is managed by staff within two of the five core departments within OSULP: the Emerging Technology and Services department and the Teaching and Engagement 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 Repositories Librarian, Digital Projects Librarian, Metadata Librarian, Digital Applications Librarian, Digital Projects Librarian, Metadata Technician, Systems Operation Technician, a team of four programmers and two student developers. A Data Management Specialist reports to the more public-facing Teaching and Engagement department, demonstrating the importance to the library of repository-related faculty and student outreach.

- The Digital Repositories Librarian (MLS, MA) 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 (MLS) is responsible for metadata associated with the two repositories, creates and manages application profiles, and ensures metadata quality and enhancement. The Metadata Technician (MLS) creates, reviews, and edits metadata for repository objects and ensures the accuracy, viability, and collection appropriateness of deposited content.
- The Digital Applications Librarian (PhD--Information Systems) assesses repository effectiveness in meeting operational library needs and develops tools to enhance repository workflows and content preservation.
- The Systems Operations Technician oversees and implements all repository systems.
- The Emerging Technologies and Services programming team is responsible for writing code in support of the repository.
- The Data Management Specialist (PhD--Oceanography) 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 Digital Projects Librarian (MLS) 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 of specimens, historical photographs, maps, journals, books, conference proceedings, creative works, and other rare and unique scholarly resources.

VI. Expert guidance

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

Compliance Level: 4 – The guideline has been fully implemented in the repository.

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/catalog?other_affiliation_label_ssim=Oregon+State+University.+Libraries>. 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 Digital Projects Librarian is 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.

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

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. Fixity checks are not regularly performed at this time.

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 are 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, a Data Management Specialist 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. As much as possible depending on the contents of the dataset, depositors are asked to provide data in non-proprietary, actionable, platform-independent formats.

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

The ScholarsArchive@OSU collection development policy 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.

Metadata technicians reserve the right to enhance and edit metadata supplied by depositors. In cases where the technician 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. A generic readme template <https://guides.library.oregonstate.edu/ld.php?content_id=45294345> is available to depositors for modification and use.

Data files must be presented, wherever possible, in formats that are non-proprietary (e.g. .csv should be used instead of .xlsx). Data 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. ScholarsArchive@OSU will accept file formats that do not comply with the above criteria 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.

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

ScholarsArchive@OSU repository policies and procedures are managed by the Digital Repository Librarian and the Metadata Technician. Policy changes or additions are posted to a public space for repository policies and, in some cases, are vetted by library administration before becoming codified <<https://wiki.library.oregonstate.edu/confluence/x/bivWAw>>.

ScholarsArchive@OSU relies on Amazon Web Services (AWS) for storage and backups, using S3 <<https://aws.amazon.com/s3/>> and Glacier <<https://aws.amazon.com/glacier/>>. 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/>>.

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

The Libraries' Digital Preservation Policy

<<https://cdss.library.oregonstate.edu/sites/default/files/osulpdigitalpreservationpolicy.pdf>> makes explicit the ScholarsArchive@OSU repository's commitment to preserving its digital resources through a comprehensive digital preservation program. Details of digital preservation implementation for specific types of digital content will be included in implementation plans that are expected to be completed in Fall 2019.

As described in the ScholarsArchive@OSU Collection Policy

<<https://wiki.library.oregonstate.edu/confluence/x/bivWAw>>, 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 and ensuring that all of the files and metadata in the repository are retrievable. ScholarsArchive@OSU is committed to providing bit-level preservation of all content. ScholarsArchive@OSU promises to ensure that the following common file formats (among many others) are retrievable in the future, using whatever combination of techniques (such as migration, emulation, etc.) is appropriate given the context of need: PDF (.pdf), XML (.xml), Text (.txt), HTML (.htm, .html), JPEG (.jpg, .jpeg), GIF (.gif), and TIFF (.tif, .tiff).

ScholarsArchive@OSU is format agnostic; 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/research-data-services/data-management-types-formats>>.

ScholarsArchive@OSU maintains a complete change history of managed content by automatically logging versions and changes associated with items in the repository. The repository relies upon a primary preservation strategy of replication of content to geographically-dispersed sites. All graduate level theses and dissertations and the University's Extension and Experiment Station Communications publications contained in the repository are preserved using the MetaArchive Cooperative distributed digital preservation system. Replication of all repository content using MetaArchive is currently in the testing phase. 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 criteria addressed:
https://metaarchive.org/wp-content/uploads/2017/03/ma_trac_audit.pdf.

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

After content is deposited to the repository it is reviewed for quality by a repository technician or the Data Management Specialist. The quality check is carried out to ensure that the data and documentation meet the requirements specified in the 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 on 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?

Several metadata elements are controlled. The Metadata Application Profile for the repository is publicly available on Google Drive <<https://docs.google.com/spreadsheets/d/1koKjV7bjn7v4r5a3gsowEimljHiAwbwuOgjHe7FEtuw/edit?usp=sharing>>. Links to related works within the repository and external to it are provided as appropriate and at the discretion of the depositor.

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

The Metadata Application Profile

<https://docs.google.com/spreadsheets/d/1koKjV7bjn7v4r5a3gsowEimljHiAwbwuOgjHe7FEtuw/edit?usp=sharing> is applied to repository deposit templates based on material type. Data are ingested to the repository by eligible users. Flexible access controls that allow depositors to restrict access to 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.

Each deposit undergoes file characterization and queuing for backup. All metadata is manually reviewed by a metadata technician 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.

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

All content housed within the ScholarsArchive@OSU repository is discoverable in Google Scholar, Google, and other internet search engines. DOIs are allocated and registered through DataCite for all datasets housed in the repository. The DOI is displayed in the DOI metadata field on dataset landing pages. ScholarsArchive@OSU persistent URLs assigned to every item in the repository are shown prominently at the top of item landing pages.

ScholarsArchive@OSU is optimized for search engine optimization (SEO) by including schema.org <http://schema.org/> JSON-LD

<https://json-ld.org/> metadata as an HTML `<script>` block on dataset landing pages, as can be verified with Google's structured data testing tool, e.g.

<https://search.google.com/structured-data/testing-tool/u/0/#url=https%3A%2F%2Ffir.library.oregonstate.edu%2Fconcern%2Fdatasets%2F6q182r738>. Work is being conducted to match all metadata elements with schema.org structured data profile. Repository metadata is harvestable using OAI-PMH for search and retrieval in the Oregon Explorer natural resources digital library <http://oregonexplorer.info/>. OSF SHARE <https://share.osf.io/> indexes

ScholarsArchive@OSU datasets for which DOIs have been registered through DataCite.

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>

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

Response: The repository exposes structured metadata for reuse via the Open Archives Initiative Protocol for Metadata Harvesting. The ScholarsArchive@OSU metadata application profile describes required and suggested metadata labels, predicates, field vocabularies and other properties for all repository work types

<<https://docs.google.com/spreadsheets/d/1koKjV7bjn7v4r5a3gsowEimljHiAwbwuOgjHe7FEtuw/edit#gid=0>>. 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. Other optional elements--funder, keywords, abstract, methodology, usage notes, related datasets/publications, and location (point, bounding box, or place name)--may also be supplied, 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. ScholarsArchive@OSU supports the full DataCite 4.0 schema <<http://schema.datacite.org/>>.

ScholarsArchive@OSU requires the creation of a documentation file(s), submitted in a non-proprietary format (e.g. .txt, .pdf) for recording and sharing data-level metadata in order to make deposited datasets usable by others. A generic readme template

<https://guides.library.oregonstate.edu/ld.php?content_id=45294345> is available to depositors for modification and use. Mandatory information required in the documentation file is: Title and abstract for the dataset; Names of creators; Name and contact information of a contact person; License or explanation of restrictions placed on the data; 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>>.

The Data Management Specialist 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. As much as possible depending on the contents of the dataset, depositors are asked to provide data in non-proprietary, actionable, platform-independent formats.

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

ScholarsArchive@OSU uses the Samvera Hyrax digital asset management solution. Hyrax is open source, community-supported

<<https://samvera.org/samvera-flexible-extensible/the-samvera-community/>> and used primarily by Libraries, Archives, and Museums. Hyrax adheres to international library metadata and other standards and to acknowledged “best practice” for design and development. The primary components of Samvera Hyrax utilized by OSULP are Fedora Commons, Solr, Blacklight, and HydraHead (a Ruby on Rails plugin and gem, respectively).

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.

The original Dublin Core metadata element set was formalized in the Internet Engineering Task Force standard RFC 5791, and discussions began about making it a standard of the (US) National Information Standards Organization (NISO). This led to the publication of ANSI/NISO Z39.85-2001 and the International Standards Organization Standard 15836-2003. The most recent updates of these standards are RFC 5791 (2010), Z39-85-2012, and ISO 15836-1:2017. Publication of a Part 2 to the ISO standard, covering several dozen properties and classes that have been added to DCMI

namespaces since 1999, is expected in 2019. The DCMI Usage Board currently serves as the maintenance agency for updating documentation for DCMI Metadata Terms (ISO 15836).

The Emerging Technologies and Services department at OSULP, responsible for maintaining the repository, has an infrastructure development plan. The plan calls for a Kubernetes cluster hosted on AWS, an Application Solr Index Fedora backend, a Background job runner, Database hosted as AWS RDS, Email service hosted as AWS, and an SES Persistent data storage in AWS S3 and Glacier for long term backups. A software inventory and system documentation are maintained by the department for internal use. Bandwidth of 10 GB has proven sufficient to meet user needs and requirements for round-the-clock connectivity and usage.

XVI. Security

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

Guidance: The repository should analyze potential threats, assess risks, and create a consistent security system. It should describe damage scenarios based on malicious actions, human error, or technical failure that pose a threat to the repository and its data, products, services, and users. It should measure the likelihood and impact of such scenarios, decide which risk levels are acceptable, and determine which measures should be taken to counter the threats to the repository and its Designated Community. This should be an ongoing process.

For this Requirement, please describe:

- Procedures and arrangements in place to provide swift recovery or backup of essential services in the event of an outage.
- Your IT security system, disaster plan, and business continuity plan; employees with roles related to security (e.g., security officers); and any risk analysis tools (e.g., DRAMBORA) you use. This Requirement describes some of the aspects generally covered by others—for example, R12 (Workflows)—and is supplementary to R9 (Documented storage procedures).