



## Assessment Information

[CoreTrustSeal Requirements 2020–2023](#)

Repository:	UC3 Merritt Curated Collection - California Digital Library
Website:	<a href="https://merritt.cdlib.org/">https://merritt.cdlib.org/</a>
Certification period:	27 March 2023 - 26 March 2026
Requirements version:	CoreTrustSeal Requirements 2020-2022

This repository is owned by: **California Digital Library**

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

## CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

### Background Information

#### Repository Type

Please provide context for your repository. You can select one or multiple options.

##### Compliance level:

Not Applicable - 0

##### Response:

- Archive
- Institutional repository
- Library
- Museum
- Publication repository
- Research project repository

#### Reviews

##### Reviewer 1:

##### Compliance level:

Not Applicable - 0

##### Comments:

##### Reviewer 2:

##### Compliance level:

Not Applicable - 0

##### Comments:

##### Reviewer 1:

##### Compliance level:

Not Applicable - 0

##### Comments:

#### Description of Repository

Provide a short overview of the repository.

##### Compliance level:

Not Applicable - 0

##### Response:

Merritt <<https://merritt.cdlib.org/>> is a general-purpose digital repository supporting curation and preservation services for the entire University of California community. It holds significant collections of digital material in all genres: arts, humanities, sciences (life, physical, and social). Historically, this content originates from UC's libraries, archives, and museums. Merritt also manages preservation copies of publications from CDL's e-Scholarship institutional repository <<https://escholarship.org/>>, the Calisphere digital collection portal <<https://calisphere.org/>>, and the archival materials from the Online Archive of California <<http://oac.cdlib.org/>>. These holdings include content from UC but also from public libraries, archives,

## UC3 Merritt Curated Collection - California Digital Library

museums, and other local cultural memory organizations from around California and the globe.

Note: While the UC3 Merritt team does offer preservation services to the Dryad Data Repository (<<https://datadryad.org/>>), the Dryad-deposited data in Merritt are not covered in this CoreTrustSeal application. Of the 378.3 TB of content being preserved in Merritt, Dryad datasets comprise 88.7 TB.

### Reviews

#### Reviewer 2:

##### Compliance level:

Not Applicable - 0

##### Comments:

#### Reviewer 1:

##### Compliance level:

Not Applicable - 0

##### Comments:

#### Reviewer 2:

##### Compliance level:

Not Applicable - 0

##### Comments:

### Designated Community

#### Provide a clear definition of the Designated Community

##### Compliance level:

Not Applicable - 0

##### Response:

Use of Merritt <<https://merritt.cdlib.org/>> is open to all faculty, students, research staff and partners of the 10-campus University of California (UC) system for contribution, discovery, and retrieval. Thus, its primary designated community encompasses scholars from all disciplines, spanning the arts, sciences, humanities, and professions. The University also has a strong public service mission and Merritt manages cultural heritage materials on behalf of local public libraries, historical societies, and other cultural memory organizations across the state of California.

Merritt collections are typically designated as restricted by their collection managers. Only a small handful of collections are made publicly available. Though the vast majority of the Merritt corpus is restricted, the material is available for retrieval through partners and/or CDL-supported discovery and access systems, such as the eScholarship institutional repository <<https://escholarship.org/>> for scholarly publications and the Calisphere portal <<https://calisphere.org/>> for archival materials. Materials may also be securely obtained through Merritt's API.

### Reviews

#### Reviewer 1:

##### Compliance level:

Not Applicable - 0

##### Comments:

#### Reviewer 2:

##### Compliance level:

Not Applicable - 0

## UC3 Merritt Curated Collection - California Digital Library

**Comments:**

**Reviewer 1:**

**Compliance level:**

Not Applicable - 0

**Comments:**

**Level of Curation**

Select all relevant types of curation.

- Content distributed as deposited
- Basic curation – e.g., brief checking, addition of basic metadata or documentation
- Enhanced curation – e.g., conversion to new formats, enhancement of documentation
- Data-level curation – as above, but with additional editing of deposited data for accuracy

**Compliance level:**

Not Applicable - 0

**Response:**

- 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

**Reviews**

**Reviewer 2:**

**Compliance level:**

Not Applicable - 0

**Comments:**

**Reviewer 1:**

**Compliance level:**

Not Applicable - 0

**Comments:**

**Reviewer 2:**

**Compliance level:**

Not Applicable - 0

**Comments:**

**Level of Curation - explanation**

Please add the description for your Level(s) of Curation.

**Compliance level:**

Not Applicable - 0

**Response:**

## UC3 Merritt Curated Collection - California Digital Library

Merritt material totals 378.3 TB and is organized into 377 curatorial collections submitted by 35 curatorial units spread across all 10 University of California campuses. Based on Producer file count, and excluding Dryad content, our material is approximately 63% text, 30% static image, 5.1% data/software, 0.6% sound, and 0.1% moving image. Merritt administrators do review, correct, and enhance contributed content as necessary.

### Reviews

#### Reviewer 1:

##### Compliance level:

Not Applicable - 0

##### Comments:

#### Reviewer 2:

##### Compliance level:

Not Applicable - 0

##### Comments:

#### Reviewer 1:

##### Compliance level:

Not Applicable - 0

##### Comments:

### Insource/Outsource Partners

If applicable, please list them.

##### Compliance level:

Not Applicable - 0

##### Response:

Merritt's external partnerships are shown in the diagram at [https://github.com/CDLUC3/mrt-doc/blob/main/img/diagrams/Merritt\\_partnerships\\_2022.pdf](https://github.com/CDLUC3/mrt-doc/blob/main/img/diagrams/Merritt_partnerships_2022.pdf). Merritt relies on other CDL services for identifier management and external service providers for storage and VM hosting. Merritt uses CDL's EZID service <http://ezid.cdlib.org> for assigning, managing, and resolving persistent ARK identifiers for all managed objects.

As shown in the diagram, Merritt incorporates a storage-broker architecture that permits it to rely upon three third-party service providers for preservation storage. One is internal to the University of California (UC) system at the San Diego Supercomputer Center (SDSC) [https://www.sdsc.edu/services/ci/universal\\_scale\\_storage.html](https://www.sdsc.edu/services/ci/universal_scale_storage.html). SDSC has not been evaluated as part of any TDR assessment, but is routinely subject to Nessus scans <http://www.tenable.com/products/nessus-vulnerability-scanner>, a professional auditing service that probes for vulnerabilities and malware. The service level agreement defining the terms of the contractual relationships between CDL and SDSC is available at [https://cdlib.org/wp-content/uploads/2020/10/SDSC\\_Service\\_Agreement\\_CDL\\_USS.pdf](https://cdlib.org/wp-content/uploads/2020/10/SDSC_Service_Agreement_CDL_USS.pdf) [UC Curation Center, 2020].

Merritt also relies on two non-UC commercial service providers – Amazon Web Services (AWS) and Wasabi Cloud Storage for preservation storage: AWS S3 <https://aws.amazon.com/s3/> and Glacier <https://aws.amazon.com/glacier/> are used for preservation storage, while database hosting is provided through use of RDS <https://aws.amazon.com/rds/> and virtual server hosting via EC2 <https://aws.amazon.com/ec2/>. Geographically speaking, these services are located on the U.S. West coast in Oregon. 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 [Amazon, 2021]. The service level agreements defining the terms of the contractual relationship between CDL and Amazon are available at <https://aws.amazon.com/agreement/> > <https://aws.amazon.com/s3/sla/> and <https://aws.amazon.com/ec2/sla/> >

Wasabi Cloud Storage <https://wasabi.com/rcs/> is used as preservation storage for an additional object copy located on the U.S. East coast in Virginia. Wasabi complies with a number of regulatory and professional IT standards <https://wasabi.com/compliance/> and certification programs including HIPAA, FERPA, SOC 2, ISO 27001 and PCI-DSS. The service level agreement that defines the terms of the contractual relationship between the University of California Office of the President and Wasabi is available at [https://cdlib.org/wp-content/uploads/2020/10/Wasabi\\_Univ\\_of\\_CA\\_Customer\\_Agreement\\_17Jan2020\\_fully-executed.pdf](https://cdlib.org/wp-content/uploads/2020/10/Wasabi_Univ_of_CA_Customer_Agreement_17Jan2020_fully-executed.pdf) [UC Curation Center, 2020].

All Merritt content is subject to Merritt-managed replication to three independent storage locations as described in response to Requirement R7.

**Reviews**

**Reviewer 2:**

**Compliance level:**

Not Applicable - 0

**Comments:**

**Reviewer 1:**

**Compliance level:**

Not Applicable - 0

**Comments:**

**Reviewer 2:**

**Compliance level:**

Not Applicable - 0

**Comments:**

**Significant Changes**

**Summary of Significant Changes Since Last Application if applicable.**

**Compliance level:**

Not Applicable - 0

**Response:**

-

**Reviews**

**Reviewer 1:**

**Compliance level:**

Not Applicable - 0

**Comments:**

**Reviewer 2:**

**Compliance level:**

Not Applicable - 0

**Comments:**

**Reviewer 1:**

**Compliance level:**

Not Applicable - 0

**Comments:**

**Other Relevant Information**

## UC3 Merritt Curated Collection - California Digital Library

You may provide other relevant information that is not covered by the requirements.

### Compliance level:

Not Applicable - 0

### Response:

Merritt material totals 378.3 TB and is organized into 377 curatorial collections submitted by 35 curatorial units spread across all 10 University of California campuses. Based on Producer file count, and excluding Dryad content, it is approximately 63% text, 30% static image, 5.1% data/software, 0.6% sound, and 0.1% moving image.

Merritt is registered with Re3data <<https://www.re3data.org/repository/r3d100010747>>.

### Reviews

#### Reviewer 2:

### Compliance level:

Not Applicable - 0

### Comments:

#### Reviewer 1:

### Compliance level:

Not Applicable - 0

### Comments:

#### Reviewer 2:

### Compliance level:

Not Applicable - 0

### Comments:

## Organizational Infrastructure

### R1 Mission/Scope

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

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Response:

The California Digital Library (CDL) exists to support the University of California (UC) community's pursuit of scholarship and extend the University's public service mission <<https://cdlib.org/about/mission-vision-and-values/>>. Within the CDL, responsibility for long-term digital curation, preservation, and research data management falls under the purview of the UC Curation Center (UC3) <<http://uc3.cdlib.org/>>. UC3 has collaborated with UC community on digital preservation projects since 2005 in furtherance of the UC Libraries' strategic goal of open scholarship to "maximize discovery of and access to information resources" in order to "ensure that the cultural and scholarly record is preserved and accessible" (University of California Libraries, 2020) <[https://libraries.universityofcalifornia.edu/wp-content/uploads/2012/11/FY20-21\\_AnnualPlansAndPriorities\\_Final.pdf](https://libraries.universityofcalifornia.edu/wp-content/uploads/2012/11/FY20-21_AnnualPlansAndPriorities_Final.pdf)> [University of California Libraries, 2020]. This is consistent with UC3's own mission "to provide transformative preservation, curation, and research data management systems, services, and initiatives that sustain and promote open scholarship" <<https://uc3.cdlib.org/who/>>, promulgated by the UC3 Program Director, CDL Executive Director, and CDL's internal leadership group (Director's Cabinet).

### Reviews

#### Reviewer 1:

## UC3 Merritt Curated Collection - California Digital Library

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

### Reviewer 2:

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

### Reviewer 1:

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

### R2 Licenses

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

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Response:

Merritt terms of use conform to the general California Digital Library (CDL) terms of use <<http://www.cdlib.org/about/terms.html>> [CDL, 2021]. Material contributed to Merritt is covered by the terms of campus-level agreements granting CDL a non-exclusive, perpetual, but revocable license to store, copy, augment, federate, and, if so curatorially-designated, distribute for non-commercial use <[https://cdlib.org/wp-content/uploads/2022/09/DPR\\_Submission\\_-Agreement\\_-8-18-06\\_FINAL.pdf](https://cdlib.org/wp-content/uploads/2022/09/DPR_Submission_-Agreement_-8-18-06_FINAL.pdf)> [CDL, 2006].

Access to Merritt content is determined by the curatorially-assigned access control rules for the collection of which the content is a member, which permit designation for either: (1) restricted access by identified individuals (authenticated via Merritt user account information or campus institutional credentials and Shibboleth-conforming IdPs); or (2) unconstrained anonymous public access and use.

The Merritt terms of use <<https://doi.org/10.24343/C3RP4S>> [UC Curation Center, 2022] state that use of Merritt constitutes acceptance by users of a commitment to abide by all applicable laws, regulations, policies, ethical concerns, and disciplinary best practices regarding the use of that content, including obligations regarding intellectual property rights, privacy, and accepted norms of scholarly discourse. All users must undergo a pre-submission consultation where these guidelines are detailed, and any user found to be exhibiting inappropriate behavior may be subject to loss of user privileges.

With regard to content removal from restricted access collections or DMCA takedown requests related to content in publicly accessible collections, the persistent identifier (ARK) assigned to an object as provided by CDL's EZID service will be altered to a tombstone status.

Merritt currently is not an appropriate repository for managing content including clinical or personally identifiable information (PII) whose disclosure would constitute a violation of HIPAA/HITECH, FERPA, or other similar statutory, regulatory, or scholarly ethical regimes. It is the contributor's responsibility to redact or anonymize content containing PII appropriately prior to submission to Merritt. Contributors may request the removal of an object if it is discovered to include such information, at which point a member of the Merritt team evaluates the request and data and removes the object from the repository.

### Reviews

### Reviewer 2:

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

### Reviewer 1:



## UC3 Merritt Curated Collection - California Digital Library

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

### Reviewer 2:

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

### R3 Continuity of access

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

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Response:

The University of California (UC) <<https://www.universityofcalifornia.edu/>> has been in existence for over 150 years, and is the nation's preeminent public institution of higher education, and among the world's best. The California Digital Library (CDL) is a core unit of the University's executive office of the systemwide President <<https://www.ucop.edu/>> with secure permanent funding. Merritt operates on a partial cost-recovery basis; while there is no service fee, data owners are assessed annually for the storage capacity that they have used. There is no defined preservation period. The longevity of ongoing preservation management is contingent on payment of associated storage costs. As detailed as part of its policies and user guidelines <<https://doi.org/10.24343/C3RP4S>> [UC Curation Center, 2022], Merritt content that is no longer being paid for is subject to review with the content owner and possible deaccessioning at CDL's discretion. In the event that CDL is unwilling or unable to continue to offer Merritt as a service to the University community, it will work with content contributors and curators to identify other curatorial institutions, within or outside the UC system, willing to take on future custodial responsibility. If that is not possible, CDL will return all content to its contributors at no added expense. Both of these contingency plans are made publicly available on the web <<https://cdlib.org/services/uc3/merritt/merritt-policies-and-procedures/>> and may be downloaded in document form at <<https://doi.org/10.24343/C3RP4S>> [UC Curation Center, 2022].

### Reviews

### Reviewer 1:

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

### Reviewer 2:

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

### Reviewer 1:

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

### R4 Confidentiality/Ethics

## UC3 Merritt Curated Collection - California Digital Library

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

### **Compliance level:**

The guideline has been fully implemented in the repository - 4

### **Response:**

The CDL is a central service unit of the University of California (UC) and the use of Merritt is open to community members on all 10 UC campuses. All content being preserved in Merritt is sourced from campus libraries, museums, UC-affiliated organizations, and our partners. Each organization is responsible for the integrity of its content and is expected to meet submission requirements.

The Merritt terms of service <<https://doi.org/10.24343/C3RP4S>> [UC Curation Center, 2022] highlight the obligation on the part of content contributors and consumers to comply with all appropriate legal, regulatory, institutional, and disciplinary requirements, policies, and ethical norms of scholarly best practice. Contribution of material to Merritt is taken as affirmation that the contributor accepts those terms of service obligations. CDL actively monitors contributions when a new collection is established through consultation and use of its collection intake forms <<https://doi.org/10.24343/C31590>>. After a collection is created and additions are made, Merritt collaborates with our partners to ensure continued compliance with this policy. Users found to be exhibiting inappropriate behavior will be subject to loss of user privileges.

Merritt is not an appropriate repository for content and 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. Merritt administrators actively remove inappropriate content as it is recognized.

In order to further improve the knowledge of UC staff (and potential Merritt collaborators) in their understanding of best practices related to digital preservation, accessibility and reuse, the UC Digital Preservation Leadership Group (DPLG) <<https://libraries.universityofcalifornia.edu/dplg>> is actively developing a formal digital preservation training program in UC's fiscal year 2022-2023.

### **Reviews**

#### **Reviewer 2:**

##### **Compliance level:**

The guideline has been fully implemented in the repository - 4

##### **Comments:**

#### **Reviewer 1:**

##### **Compliance level:**

The guideline has been fully implemented in the repository - 4

##### **Comments:**

#### **Reviewer 2:**

##### **Compliance level:**

The guideline has been fully implemented in the repository - 4

##### **Comments:**

### **R5 Organizational infrastructure**

**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:**

The guideline has been fully implemented in the repository - 4

### **Response:**

The California Digital Library (CDL) is a central administrative unit of the University of California (UC) <<https://www.universityofcalifornia.edu/>>, one of the world's premier public universities. UC was founded in 1869 and now hosts over 285,000 students and 167,000 faculty and staff on 10 campuses, five

## UC3 Merritt Curated Collection - California Digital Library

medical centers, and three national laboratories, with an annual budget of \$39.8 billion (FY 2019-2020), supported by federal and state government appropriations, public and private grants and contracts, endowment, and tuition and fees  
<[https://www.ucop.edu/operating-budget/\\_files/rbudget/2020-21-budget-summary.pdf](https://www.ucop.edu/operating-budget/_files/rbudget/2020-21-budget-summary.pdf)> [University of California, 2020]. The CDL, founded in 1997, has an annual budget of \$27.8M. CDL budgeting is performed on an annual basis.

Merritt has no service fees, but Merritt users are billed for the recovery cost of preservation storage <<https://doi.org/10.24343/C3RP4S>> [UC Curation Center, 2022]. Billing rates are set at the cost of provisioning storage by CDL's external storage providers at the San Diego Supercomputer Center (SDSC), Amazon AWS and Wasabi, with an added 4% contingency surcharge to build up modest surpluses to be used in event of non-periodic or unanticipated expenses.

Merritt is supported by the UC Curation Center (UC3), one of four core programmatic units at the CDL. The UC3 team includes permanent, full-time, dedicated roles for a Digital Preservation Services Manager, a development manager, frontend and backend programmers, and a DevOps/system administrator. UC3 also has access to the services of CDL's User Experience (UX) assessment and design team (6 FTE) <<http://www.cdlib.org/services/uxdesign/>> and Infrastructure and Application Support (IAS) team (4 FTE) <<http://www.cdlib.org/services/infrastructure/>>. CDL staff are widely recognized as international experts in their fields, and routinely participate in professional organizations, activities, and conferences, including NDSA DigiPres <<https://ndsa.org/conference/>>, Open Repositories <<https://www.openrepositories.org/>>, iPRES <<https://ipres-conference.org/>>, Future of Research Communications and e-Scholarship (FORCE11) <<https://www.force11.org/>>, and Data Carpentry (DC) <<http://www.datacarpentry.org/>> and Library Carpentry (LC) <<https://librarycarpentry.org/>>.

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

### R6 Expert guidance

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

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Response:

Within the UC Office of the President, CDL falls under the purview of the vice provost for Academic Personnel and Programs. Systemwide accountability for CDL's operation is provided through frequent reporting to, and consultation with, the Council of University Librarians (CoUL), consisting of the 10 campus ULs and the CDL executive director <<https://libraries.universityofcalifornia.edu/coul>>, who meet monthly to discuss issues of systemwide policy and initiatives, and the Systemwide Library and Scholarly Information Advisory Committee (SLASAC), consisting of representatives drawn from systemwide vice chancellors, vice provosts, CIOs, faculty, and the CDL executive director <<https://libraries.universityofcalifornia.edu/slasiac>>.

The CDL receives general guidance on its activities from the UC Libraries Advisory Structure (UCLAS) <<https://libraries.universityofcalifornia.edu/about/advisory-structure>> and its Direction and Oversight Committee (DOC) <<https://libraries.universityofcalifornia.edu/doc>>. It is a participant in and gains guidance from the Digital Preservation Leadership Group (DPLG), whose charge is to lead the University of California in the area of UC-wide digital preservation

## UC3 Merritt Curated Collection - California Digital Library

<<https://wiki.library.ucsf.edu/pages/UC+Digital+Preservation+Leadership+Group>>. CDL also relies on a number of systemwide Common Knowledge Groups (CKG) with broad campus participation by local library and RDM staff for more specific consultation and guidance, including the Born-Digital CKG <<https://wiki.library.ucsf.edu/display/UCLCKG/Born+Digital+Planning+Group+CKG>>, Data Curation CKG <<https://wiki.library.ucsf.edu/display/UCLCKG/Data+Curation+CKG>>, and Digital Repository Metadata CKG <<https://wiki.library.ucsf.edu/display/UCLCKG/Digital+Repository+Metadata+CKG>>.

The CDL works closely with campus-based data librarians and data curation programs, particularly the Research Data Management group at UC Berkeley <<http://researchdata.berkeley.edu/>>, in which the CDL holds an advisory position, the UC Davis Data Science Initiative <<https://datalab.ucdavis.edu/data-science-initiative/>>, the UC Santa Barbara data curation group <<https://www.library.ucsb.edu/data-curation>>, and the UC San Diego research data curation program <<https://library.ucsd.edu/research-and-collections/data-curation/>>.

The CDL holds institutional memberships in the Coalition for Networked Information (CNI) <<http://www.cni.org/>>, Council on Library and Information Resources <[www.clir.org](http://www.clir.org)>, COUNTER <<http://www.projectcounter.org/>>, DataCite <<http://www.datacite.org/>>, Digital Library Federation (DLF) <<http://www.diglib.org/>>, EDUCAUSE <[www.educause.edu](http://www.educause.edu)>, HathiTrust <[www.hathitrust.org](http://www.hathitrust.org)>, International Coalition of Library Consortia (ICOLC) <[icolc.net](http://icolc.net)>, International Internet Preservation Consortium (IIPC) <<http://netpreserve.org/>>, National Digital Stewardship Alliance (NDSA) <<http://nds.org/>>, National Information Standards Organization (NISO) <<http://www.niso.org/>>, OCLC <<https://www.oclc.org/en/home.html>>, ORCID <<http://www.orcid.org/>>, Crossref <[www.crossref.org](http://www.crossref.org)>, and Scholarly Publishing Academic Resources Coalition (SPARC) <<http://sparcopen.org/>>.

The Digital Preservation Services Manager routinely communicates with designated communities through targeted email to the systemwide CDLINFO distribution list <<https://www.cdlib.org/cdlinfo/>>, email lists of campus collection collaborators, and posts to the UC3 blog <<http://uc3.cdlib.org/>>.

### Reviews

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

### Digital Object Management

#### R7 Data integrity and authenticity

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

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Response:

Content may be submitted to Merritt with optional cryptographic message digest values for individual file-level components <<https://github.com/CDLUC3/mrt-ingest/blob/main/README.md>>. If present, these digests are verified as part of ingest processing of Submission Information Packages (SIP). If individual files are not already associated with SHA-256 message digests, they are automatically assigned during the creation of the Archival Information Package (AIP). These digests are automatically verified following the transfer of the AIP from the ingest process to the archival storage process, by an immediate readback. The digests are also verified on retrieval of the AIP for external distribution as a Dissemination Information Package (DIP).

## UC3 Merritt Curated Collection - California Digital Library

All content managed in Merritt is replicated to three, distributed storage locations relying on heterogeneous technologies and media <<https://github.com/CDLUC3/mrt-replic/blob/master/README.md> >. A multi-threaded Audit process runs continually to verify the SHA-256 digests of all dispersed replicas <<https://github.com/CDLUC3/mrt-audit/blob/master/README.md> >. Canonical digest values for all versions of all files are stored in Merritt's central Inventory database <<https://github.com/CDLUC3/mrt-inventory/blob/master/README.md> >, which runs in a mirrored, cross-availability zone configuration in the Amazon AWS RDS cloud. Over the course of the past ten years of operation, bit-level damage was identified on less than half a dozen occasions and in all cases the damage was fully corrected by copying from verified replicas.

Individual replication sites also implement internal mechanisms for ensuring integrity and authenticity. Merritt relies on SDSC's UC private cloud service, Amazon's AWS S3 and Glacier commercial cloud services and the Wasabi commercial cloud service for preservation storage. These are federated through Merritt's storage broker architecture <<https://github.com/CDLUC3/mrt-store/blob/master/README.md> >.

SDSC Universal Scale Storage <[https://www.sdsc.edu/services/ci/universal\\_scale\\_storage.html](https://www.sdsc.edu/services/ci/universal_scale_storage.html) > implements a Qumulo QF2 file storage system that provides durability by distributing erasure coding stripes across multiple storage servers. The system continuously confirms the underlying media with an ongoing process that performs verification of the disk sectors. AWS S3 and Glacier cloud storage services all maintain internal replicas and are designed to provide eleven 9's of reliability (99.99999999%) and sustain the simultaneous failure of any two internal copies. Wasabi is designed to provide a similar 99.99999999% degree of durability, and akin to Qumulo, is based on use of erasure coding algorithms that distribute object data across independent disk drives in separate storage servers. Additionally, Wasabi provides checksum comparisons on all objects every 90 days. Merritt's Audit process provides external verification of all content independent of local internal methods at SDSC, S3 and Wasabi. Glacier storage is not externally validated, however, as the AWS transactional pricing structure makes this financially prohibitive. All content stored in Glacier is also replicated at SDSC, which is subject to external validation.

Merritt is a strongly-versioned repository. Any changes to data or metadata automatically results in the creation of a new version of the data object. Versioning relies on file-level backwards deltas to minimize duplicative file storage. Individual file-level components are never edited or replaced; new versions of files are added as components of the new version. All previous versions can be retrieved through the Merritt UI and API. Relationships between an object's various versions and files are maintained in Merritt's Inventory database <<https://github.com/CDLUC3/mrt-inventory/blob/master/README.md> >. Version provenance information – effectuating agent, date/time – is also stored in the database as well as being expressed in metadata files that form part of the object's archival (AIP) and dissemination (DIP) information packages. The database also holds object-specific metadata as well as Electronic Resource Citation (ERC) metadata <<http://dublincore.org/groups/kernel/> > for all objects. These metadata are also expressed in file form as part of the object's AIP and DIP.

Submission of new content to Merritt non-public collections requires prior authenticated login using Merritt user account credentials, which are managed in a three-node LDAP high-availability cluster <<https://cdluc3.github.io/mrt-doc/diagrams/ldap> >.

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

### R8 Appraisal

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

##### Compliance level:

The guideline has been fully implemented in the repository - 4

## UC3 Merritt Curated Collection - California Digital Library

### Response:

By explicit policy <<https://doi.org/10.24343/C3RP4S>> [UC Curation Center, 2022], Merritt can accept new data in any genre, format, or structure. ■■■The Merritt Team evaluates pre-deposit the intended nature and quality of the digital objects to be ingested, the metadata strategy to be taken by the depositor, and the understanding and commitment to comply with Merritt policies and guidelines.

CDL provides detailed guidelines for recommended formats of content submitted to Merritt <<https://doi.org/10.24343/C3MW2V>> [CDL, 2022]. Merritt's submission interface provides the opportunity to supply Dublin Kernel metadata serialized in the form of an Electronic Record Citation (ERC) <<http://dublincore.org/groups/kernel/>>, with elements for creator(s), title, date, and identifier(s). Additional metadata files may be freely included as part of the data package contributed to Merritt. Responsibility for developing appropriate data creation, acquisition, and description protocols and workflows is held by the Merritt team, the professional campus librarians and/or collections managers.

CDL offers consultation and guidance on ways to acquire or create digital content in a manner that is most amenable to the highest level of future preservation. At a minimum, CDL is committed to providing bit-level preservation of all content. However, CDL's preservation policy <<https://doi.org/10.24343/C3RP4S>> [UC Curation Center, 2022] also obligates it to make reasonable efforts to provide managed content with the highest level of preservation assurance that is consistent with the form, structure, and packaging of the content, the degree to which that it is accompanied by authoritative and comprehensive metadata, the availability of appropriate tools, and other organizational priorities.

### Reviews

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

### R9 Documented storage procedures

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

##### Compliance level:

The guideline has been fully implemented in the repository - 4

### Response:

Merritt's preservation policy is outlined in its Policies and User Guidelines statement <<https://doi.org/10.24343/C3RP4S>> [UC Curation Center, 2022]. The Merritt repository is implemented in terms of a microservices architecture <<https://doi.org/10.2218/ijdc.v5i1.151>> [Abrams et al., 2010]. The overall interoperation of the various microservices are indicated in this architectural diagram <<https://github.com/CDLUC3/mrt-doc/blob/main/README.md#core-microservices>>. Merritt relies on a primary strategy of replication to ensure the long-term integrity of managed data. All data is replicated to at least three geographically distributed locations and three heterogeneous technology stacks, currently at the San Diego Supercomputer Center (SDSC) <[https://www.sdsc.edu/services/ci/universal\\_scale\\_storage.html](https://www.sdsc.edu/services/ci/universal_scale_storage.html)>, which uses Qumulo QF2 storage, Amazon AWS, which uses S3 and Glacier, and Wasabi's cloud storage implementation. The SDSC Universal Scale Storage service implements a Qumulo QF2 file storage system that provides durability by distributing erasure coding stripes across multiple storage servers. The system continuously confirms the underlying media with an ongoing process that performs verification of the disk sectors. The S3 service description <<https://aws.amazon.com/s3/>> strongly implies that it relies upon three independent replicas spread across availability zones with attendant internal fixity auditing and self-healing, and claims a 99.999999999% degree of durability. Wasabi claims a similar 99.999999999% degree of durability based on use

## UC3 Merritt Curated Collection - California Digital Library

of industry-proven erasure coding algorithms that distribute object data across independent disk drives in separate storage servers. Additionally, Wasabi provides checksum comparisons on all objects every 90 days <<https://wasabi.com/wp-content/uploads/2019/10/Durability-Tech-Brief-2019.pdf>>.

The process of accepting SIPs and transforming them into AIPs for archival management is documented in terms of the individual microservices implicated in that processing. Flow diagrams are available for each – Ingest <<https://github.com/CDLUC3/mrt-ingest/blob/main/README.md>>, which transforms the SIP into a conforming AIP; Storage <<https://github.com/CDLUC3/mrt-store/blob/master/README.md>>, which dispositions the AIP to its final primary storage location at either the AWS S3 cloud or the San Diego Supercomputer Center (SDSC); Inventory <<https://github.com/CDLUC3/mrt-inventory/blob/master/README.md>>, which retrieves newly acquired AIPs from Storage for parsing and populating the central metadata catalog and search index; Replication <<https://github.com/CDLUC3/mrt-replic/blob/master/README.md>>, which copies the AIP to its secondary storage location at the Wasabi cloud or the AWS Glacier cloud; and Audit <<https://github.com/CDLUC3/mrt-audit/blob/master/README.md>>, which registers cryptographically-secure SHA-256 message digests for all replicas of all file-level API components, and subjects them to routine periodic verification.

Reports of any discrepancies in the digests of stored replicas are generated automatically for Merritt administrators to investigate, triage, and if necessary, intervene. A reported discrepancy may not indicate actual bit-level damage; the temporary loss of online access to the remote cloud service providers is a legitimate trigger of nightly reports to administrators. The Merritt repository has been in operation since October 2010 without any data loss, and as of September 2022 manages over 3.42 million digital objects, represented in 5.43 million discrete versions and 67.7 million individual files totaling 378.3 TB. During that time, actual bit-level damage was identified on less than half a dozen occasions and in all cases the damage was fully corrected by copying from verified replicas.

### Reviews

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

### R10 Preservation plan

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

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Response:

The Merritt terms of service <<https://doi.org/10.24343/C3RP4S>> [UC Curation Center, 2022] make clear that the CDL is obligated to make reasonable efforts to provide managed content with the highest level of preservation assurance that is consistent with the form, structure, and packaging of the content, the degree to which that it is accompanied by authoritative and comprehensive metadata, the availability of appropriate tools, and other organizational priorities. There are no formal preservation "levels" per se; rather, CDL assumes a continuum of preservation outcomes dependent upon the nature of the content. At a minimum, CDL is committed to providing bit-level preservation of all content in accordance with its preservation strategy <<https://doi.org/10.24343/C3RP4S>> [UC Curation Center, 2022]. CDL offers consultation and guidance on ways to acquire or create digital content in a manner that is most amenable to the highest level of future preservation service. As collections are established and deposits made, Merritt continues to collaborate with depositors through review of content types and metadata that comprise digital objects. Merritt provides for automated classification of file types (specifically, MIME types) on ingest, the results of which are made available through repository reporting tools for evaluation at any time.

## UC3 Merritt Curated Collection - California Digital Library

Contribution to Merritt is taken as affirmative assent that the contributors are assigning to CDL the non-exclusive, perpetual, revocable right to save, copy, enhance, federate, create derivatives for purposes of long-term preservation, and provide access to contributed content, subject to curatorially-designated controls.

Merritt maintains a complete change history of managed content as it may evolve over time. The repository relies upon a primary preservation strategy of replication of content to geographically-dispersed sites and technological heterogeneity. Merritt incorporates a process of continual verification of cryptographic message digests of all content replicas to detect and correct any bit-level damage.

The design, implementation, and operation of Merritt are consistent with the community-accepted standard ISO 14721 Open Archival Information System (OAIS) reference model [ISO 14721, 2012].

### Reviews

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

A search in the preservation policy for the terms migration or curation does not return any results. The preservation policy states: "The primary preservation strategy the Merritt repository employs is bit-level preservation." I suggest you plan to add content with respect to migration and curation in your preservation policy and make it more explicit in your submission in 3 years.

#### Reviewer 1:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

#### Reviewer 2:

##### Compliance level:

The guideline has been fully implemented in the repository - 4

##### Comments:

### R11 Data quality

**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:

The guideline has been fully implemented in the repository - 4

##### Response:

The technical infrastructure of the repository combined with the expertise of the Merritt team and that of the digital preservation professionals across the University of California ensures the completeness and quality of the data and metadata of all deposits. The cornerstone of our process is the use of collection intake forms <<https://doi.org/10.24343/C31590>>. These forms allow for a consultative approach to the creation of every new collection. They provide a means to evaluate up-front the intended nature and quality of the digital objects to be ingested, the metadata strategy to be taken by the depositor and the understanding and commitment to comply with Merritt policies and guidelines.

For all deposits into Merritt, regardless of their origin (network attached storage, Digital Asset Management System, etc.), our campus-based colleagues, librarians, and/or collections managers are responsible for working with Merritt staff to ensure digital object composition and metadata quality meet all institutional retrieval requirements.

### Reviews

#### Reviewer 1:



## UC3 Merritt Curated Collection - California Digital Library

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

### Reviewer 2:

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

### Reviewer 1:

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

### R12 Workflows

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

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Response:

The nominal archiving workflow <<https://github.com/CDLUC3/mrt-doc/blob/main/README.md#core-microservices>> for Merritt starts with transfer of one or more data files to the Merritt Ingest service <<https://github.com/CDLUC3/mrt-ingest/blob/main/README.md>>, which can occur through its online UI, REST API, or SWORD v2 protocol. Ingest processing is performed by the iterative invocation of individual ingest handlers, each focusing on a particular aspect of processing, including container disaggregation, digest verification, characterization, identifier assignment, and SIP-to-AIP translation. Ingest places messages on an asynchronous queue alerting the Storage and Inventory services of the availability of new content. The Storage service <<https://github.com/CDLUC3/mrt-store/blob/master/README.md>> moves the content from a temporary staging space into its final primary storage location, either SDSC or AWS S3, under the direction of an ingest profile in which context the submission was made. The Inventory service <<https://github.com/CDLUC3/mrt-inventory/blob/master/README.md>> retrieves the content from Storage and parses its metadata files for inclusion in the Inventory database and search index. Part of the database representation includes a specification of the replication policy and fixity information. The Replication service <<https://github.com/CDLUC3/mrt-replic/blob/master/README.md>> runs continually and upon recognizing that the new content has not yet been replicated, will do so to its secondary location, either Wasabi or AWS Glacier, again dependent upon the controlling submission profile. The Audit service <<https://github.com/CDLUC3/mrt-audit/blob/master/README.md>> also runs continually, cycling through all registered content and retrieving and verifying SHA-256 message digests for all files of all versions of all replicas. Reports of any discrepancies in the digests of stored replicas are generated automatically for Merritt administrators to investigate, triage, and if necessary, intervene. All Merritt collections publish Atom feeds with the newest content on the first page. Feed metadata includes actionable links to managed content's landing pages. Content is also accessible through browsing or metadata search over contributor(s), title, and identifier metadata.

All code for Merritt services is managed in versioned-controlled git repositories on GitHub. Maintenance and enhancement of the codebases relies on documented agile development practices with a user-centered design focus. A dedicated product development team gathers stakeholder needs and stories and translates them into prioritized use cases, which are broken down into individual tickets managed in GitHub Issues. Work on the prioritized backlog is performed in three week time-boxed sprints, with incremental progress logged in the tickets.

### Reviews

### Reviewer 2:

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Comments:

### Reviewer 1:

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:****Reviewer 2:****Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:****R13 Data discovery and identification**

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

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Response:**

All objects managed in Merritt are assigned unique, persistent ARK identifiers <<https://tools.ietf.org/html/draft-kunze-ark-18> > [Kunze and Rodgers, 2013] using CDL's EZID service <<http://ezid.cdlib.org/>>.

All Merritt object landing pages prominently display the object's actionable persistent identifier(s) for use in citations. In cases where the depositor also has recommended citation practices are made available through their associated websites, those citations and instructions are detailed directly on partner websites.

Merritt provides search within collections based on indexed Dublin Kernel/ERC metadata: creators(s), title, date, and primary ARK identifier.

Merritt defines stable URL patterns for object landing pages, version landing pages, object downloads, version downloads, and file downloads of the general form:

<https://merritt.cdlib.org/m/<objectid>>

<https://merritt.cdlib.org/m/<objectid>/<versionid>>

<https://merritt.cdlib.org/api/assemble-obj/<objectid>>

<https://merritt.cdlib.org/api/assemble-obj/<objectid>/<versionid>>

<https://merritt.cdlib.org/api/presign-file/<objectid>/<versionid>/<fileid>>

where <objectid>s are ARKs, <versionid>s are ordinal numbers 1,2,3..., where 0 is a shortcut reference to the current version, and <fileid>s are pathnames (with colons ":" and slashes "/" URL-encoded). For example:

<https://merritt.cdlib.org/m/ark%253A%252F13030%252Fm52f7p63>

<https://merritt.cdlib.org/m/ark%253A%252F13030%252Fm52f7p63/3>

<https://merritt.cdlib.org/api/assemble-obj/ark%253A%252F13030%252Fm52f7p63/>

<https://merritt.cdlib.org/api/assemble-obj/ark%253A%252F13030%252Fm52f7p63/3>

<https://merritt.cdlib.org/api/presign-file/ark%253A%252F13030%252Fm52f7p63/3/producer%252FUC3-Merritt-preservation-latest.pdf>

are the URLs for an object landing page, version 3 landing page, object download, version 3 download, and main PDF file download.

Object retrieval involves an asynchronous process, where the object is first constructed and saved to a temporary cloud storage location. It is then delivered to the client through use of a presigned URL

<<https://github.com/CDLUC3/mrt-doc/wiki/Retrieving-Objects#programmatic-retrieval-of-merritt-objects> >.

Merritt ARKs (of the general form: <http://n2t.net/<ark>>) resolve to the Merritt object landing pages. For example:

<http://n2t.net/ark:/13030/m52f7p63>

is the URL for the same object landing page.

Curatorially-designated content collections syndicate their holdings for automated harvesting via ATOM feeds accessible through a prominent button on collection landing pages. This function may be used by external value-added services building aggregated cross-repository collections and discovery mechanisms.

**Reviews****Reviewer 1:****Compliance level:**

## UC3 Merritt Curated Collection - California Digital Library

The guideline has been fully implemented in the repository - 4

**Comments:**

**Reviewer 2:**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:**

**Reviewer 1:**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:**

### R14 Data reuse

**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:**

The guideline has been fully implemented in the repository - 4

**Response:**

CDL provides general guidelines for recommended formats of content and metadata submitted to Merritt <<https://doi.org/10.24343/C3MW2V>> [CDL, 2022]. Merritt's submission interface provides the opportunity to supply Dublin Kernel metadata serialized in the form of an Electronic Record Citation (ERC) <<http://dublincore.org/groups/kernel/>>, with optional elements for creator(s), title, date, and identifier(s). Additional arbitrary metadata, including domain-specific, may be freely incorporated as part of submitted Merritt object packages. As collections are established and deposits made, Merritt continues to collaborate with depositors through review of and possible enhancement to object metadata. The addition of new metadata to existing objects is supported through object versioning, which in turn enables a complete history of metadata changes to be captured over time.

Merritt's preservation policy <<https://doi.org/10.24343/C3RP4S>> [UC Curation Center, 2022] does not prescribe any particular strategy for the reuse of the digital objects it houses, but rather believes said strategy must be carefully tailored to specific use cases and designated communities.

### Reviews

**Reviewer 2:**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:**

**Reviewer 1:**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:**

**Reviewer 2:**

**Compliance level:**

The guideline has been fully implemented in the repository - 4

**Comments:**

## Technology

### R15 Technical infrastructure

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:

The guideline has been fully implemented in the repository - 4

#### Response:

Merritt is hosted on Linux VMs in the Amazon AWS EC2 cloud environment. Their associated MySQL databases are hosted by AWS RDS. General operating principles for CDL's AWS usage is described at <<https://cdlib.org/wp-content/uploads/2022/02/CUG-CDLAWSEnvironmentOrientation.pdf>>, <<https://cdlib.org/wp-content/uploads/2022/02/CUG-RDSUserGuide.pdf>>, and <<https://cdlib.org/wp-content/uploads/2022/02/CUG-S3andGlacierStorageUsersGuide.pdf>> [CDL, 2021, CDL, 2021; CDL, 2018]. Consistent with CDL policy, all VMs are routinely updated to the latest stable AWS operating systems versions within six months of release, or within one week for critical security patches or 30 days for non-critical security patches.

Merritt is available on a nominal 24x7x52 high-availability schedule. All backend Merritt processing is fully load-balanced so routine maintenance can occur without interrupting service availability. Maintenance on the one singleton process, the frontend load balancer, takes place during a published after-hours maintenance window when web traffic tends to be at its lowest. The current status of Merritt availability can be found on the CDL system status page <<http://www.cdlib.org/contact/system.html>>.

Merritt fully implements the Dublin Kernel/Electronic Record Citation (ERC) community metadata standard <<http://dublincore.org/groups/kernel/>>, OAI-PMH 2.0 metadata harvesting standard, and SWORD v2.

Merritt incorporates the community-supported open source Jersey REST framework <<https://eclipse-ee4j.github.io/jersey/>> and Zookeeper coordination suite <<https://zookeeper.apache.org/>>. Individual Merritt microservices are Java applications running on Apache Tomcat application servers <<http://tomcat.apache.org/>>; their source code and documentation are managed under revision control in public GitHub repositories <<https://github.com/CDLUC3>> and <<https://github.com/CDLUC3/mrt-doc>>.

Merritt infrastructure configuration parameters are stored in an AWS Systems Manager Parameter store. Configuration management and software deployments pertaining to Merritt microservice EC2 hosts are enabled through Puppet automation <<https://puppet.com/>>. The location where Puppet code resides is a restricted access, private git repository.

In order to track characteristics of EC2 hosts, an AWS Lambda function-based system inventory application has been created and is accessible only to CDL personnel through VPN connectivity. Server certificates are managed centrally by the CDL Infrastructure and Application Support (IAS) program.

## Reviews

### Reviewer 1:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

### Reviewer 2:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

### Reviewer 1:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

## R16 Security

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

### Compliance level:

The guideline has been fully implemented in the repository - 4

### Response:

Merritt is deployed and operates on virtual machines in the Amazon AWS cloud. AWS is certified for ISO 27001 [Amazon, 2016b] and BSI C5 <<https://aws.amazon.com/compliance/bsi-c5/>> security standards. AWS's comprehensive security controls, including those for both online and physical facility security, are described at <<https://aws.amazon.com/compliance/data-center/controls/>>.

All Merritt processes run on standard Amazon Linux VMs in the context of service-specific role accounts that are accessible only by individual UC3 staff members through explicitly-granted sudo privileges (sudo su - <role>). All grants of sudo privileges are logged and monitored by CDL system administrators. Personal login accounts for Merritt hosts are limited to appropriate CDL staff and are accepted only via ssh from the IP range of CDL's administrative office, its virtual private network (VPN), or by ssh from a hardened bastion server.

Software deployments to Merritt microservice hosts are controlled automatically using Puppet operating in a master/agent configuration <<https://puppet.com/>>. Nagios <<https://www.nagios.org/>> is used for continuous monitoring of VM resource levels for disk space, load average, and swap usage. Additional Nagios monitors are defined for key service functions to provide early notification in case of interruptions to online availability.

All AWS EC2 virtual servers are backed up nightly with a full instance snapshot retained for 35 days, with the first snapshot of each month retained for six months. Weekly machine image snapshots, including all data volumes and instance configuration information, are maintained outside of the primary AWS region. CDL central IT maintains backups of all AWS configuration information. Many resources, including IAM, security groups, S3 bucket configurations, Route 53 DNS records, and CloudFormation templates are under version control in a local git repository (itself backed up in an alternate region)

<<https://cdlib.org/wp-content/uploads/2022/02/CUG-CDLAWSEnvironmentOrientation.pdf>> [CDL, 2021]. MySQL databases, configured for mirrored operation in multiple RDS availability zones, are also backed up nightly and are retained for 35 days

<<https://cdlib.org/wp-content/uploads/2022/02/CUG-S3andGlacierStorageUsersGuide.pdf>> [CDL, 2018]. CDL-wide disaster recovery plans call for redeployment of services from image snapshots in the primary or another AWS region as dictated by necessity.

All CDL staff participate in mandatory annual cybersecurity training <<https://security.ucop.edu/>> [UC Office of the President, 2021].

## Reviews

### Reviewer 2:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

### Reviewer 1:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

### Reviewer 2:

#### Compliance level:

The guideline has been fully implemented in the repository - 4

#### Comments:

## Applicant Feedback

### R17 Applicant Feedback

We welcome feedback on the CoreTrustSeal Requirements and the Certification procedure.

#### Compliance level:

## UC3 Merritt Curated Collection - California Digital Library

The guideline has been fully implemented in the repository - 4

### **Response:**

Given this is our first recertification, CDL and the Merritt team appreciate the opportunity to discuss the progress that has been made with the repository over the past three years. As with our prior submission, this one has been an excellent exercise for the team to undertake in terms of reviewing policies and workflows, and summing up our efforts associated with the implementation of new functionality and the application of technology.

### **Reviews**

#### **Reviewer 1:**

##### **Compliance level:**

The guideline has been fully implemented in the repository - 4

##### **Comments:**

These modifications are acceptable and I recommend certification at this time.

#### **Reviewer 2:**

##### **Compliance level:**

The guideline has been fully implemented in the repository - 4

##### **Comments:**

Suggestion for improvement for resubmission in 3 years: Document how you approach and take appropriate actions to ensure that the quality of metadata is "sufficient for continued understandability of the preserved content by the Designated Community" (CTS Extended Guidance, p. 23).