

Merritt Digital Preservation Repository

Policies and User Guidelines

Version 2.1

Purpose

These policies and user guidelines describe the library intent and strategies associated with the [California Digital Library](#) (CDL, or “the Library”) [Merritt](#) digital preservation repository. They provide context surrounding the repository, outline CDL and contributor responsibilities, review the systematic application of technology to provide preservation assurance, and include guiding information associated with the ingest, preservation and stewardship of content over time.

Contents

Part I: Context and Services

Part II: Digital Preservation Strategy

Part III: Privacy, Accessibility and Responsibilities

Part IV: Format Guidelines, Content Versioning and Persistent Identifiers

Part V: Service Providers and Storage Costs

Part VI: Appendices

Part I: Context and Services

The California Digital Library exists to support the University of California community’s pursuit of scholarship and extend the University’s public service mission. The Merritt digital preservation

California Digital Library

repository is a core CDL service available for use by all members of the UC community for managing, preserving, publishing, and sharing the University's valuable digital content.

Merritt complies with the general [CDL terms of service](#) as well as the terms presented within this policy guide.

Part II: Digital Preservation Strategy

Digital preservation is a combination of actors, institutional policies, procedures and technologies that are all geared to ensure access to digitized and born digital content over the course of time, regardless of change in any one of these elements. In the context of digital preservation, the notion of providing *access* refers to ensuring the continuity of content *usability*, *authenticity* and *integrity*.

CDL offers consultation and guidance on ways to acquire or create such digital content in a manner that is most amenable to the highest level of future preservation service. As collections are established and deposits made into the system, Merritt collaborates with depositors through review of content types and metadata that comprise digital objects. In this vein, Merritt provides for automated classification of file types on ingest, the results of which are made available through repository reporting tools for evaluation at any time.

The primary preservation strategy the Merritt repository employs is *bit-level* preservation. Preservation at the bit level is purposed to safeguard the bits of each file stored in the repository – bits being the series of 1s and 0s that encode the meaning of the digital materials they form. Success in bit-level preservation sees each bit of every file remain unchanged, or “fixed” over time.

To fortify its strategy, Merritt actively manages three copies of all files and digital objects in the system through use of external storage providers for primary and replication storage. The content of all collections in Merritt benefits from three object copies, maintained across three different cloud storage providers distributed across two geographic regions (US West Coast, and US East Coast) with differing disaster threats in order to mitigate risk.

Content in Merritt is organized into collections. A collection is composed of one or more digital objects, each object containing a series of digital files. Through object versioning, Merritt maintains a complete change history of managed content as it may evolve over time. All files are routinely

California Digital Library

fixity-checked through continual verification of cryptographic message digests of all content replicas to detect and correct any bit-level damage. Fixity checking cycles are completed across the entire corpus within a period of 90 days or less. Errors with ingest, replication, inventory, or storage operations are reported through automated system consistency checks which run on a daily basis.

The design, implementation, and operation of Merritt are consistent with the community-accepted standard [ISO 14721](#) Open Archive Information System (OAIS) reference model. In addition, CDL aims to provide enhanced services by offering initial and ongoing expert consultation and guidance on ways to acquire or create digital content and metadata in a manner that is most amenable to the highest level of future preservation service.

Part III: Privacy, Accessibility and Responsibilities

PRIVACY

Merritt complies with [CDL's privacy policy](#), under which the privacy of all users will be respected and protected in compliance with federal and state laws and [University of California Policies](#).

ACCESSIBILITY

As part of the University of California, CDL is required to follow the [University of California Information Technology Accessibility Policy](#) to help ensure that as broad a population as possible may access, benefit from, and contribute to the University's electronic programs and services. The Policy requires that all web content in Merritt complies with WCAG 2.1 Guidelines at level AA success criteria. Accordingly, contributors are responsible for ensuring that content they deposit to Merritt meets this standard.

CONTRIBUTOR RESPONSIBILITIES

By contributing to Merritt, content owners are acknowledging that they have followed all applicable laws, regulations, policies, ethical concerns, and disciplinary best practices regarding the creation and acquisition of that content, including obligations regarding intellectual property rights, privacy, IRB review, and accepted norms of scholarly discourse, and that they assign 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 access controls for the collection of which the content is a member. Said controls permit designation for

California Digital Library

either authenticated access and use only by a restricted set of individuals, or unconstrained public access and use. Contributors exhibiting inappropriate behavior will be subject to loss of user privileges.

Merritt 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 ethical regimes. Content containing PII must, however, be redacted or anonymized prior to submission to Merritt.

CDL RESPONSIBILITIES

CDL accepts, manages, and provides access to digital content in order to support the University's research, teaching, learning, and public service mission. CDL will not exploit managed content in profit-generating activity without express permission of its legal owners.

CDL makes 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. Note that this implies 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. However, CDL aims to provide enhanced services by offering initial and ongoing expert 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.

Merritt relies on browser-based cookies to maintain online session information for streamlining user experience. All access log information and other personally-identifying evidence of use is collected and dispositioned in a manner consistent with the CDL privacy policy.

In the event that CDL is unable or unwilling to continue operation of Merritt, it will make reasonable efforts to find another curatorial organization, within or outside the UC system, willing to take on custodial responsibility for all managed content. If that is not possible, CDL will return all content to its contributors at no added expense.

DEACCESSIONING

Merritt is operated on a partial cost-recovery basis, as described in *Service Providers and Storage Costs*. At any time, contributors may request a bulk export of their content, for which CDL may impose a

California Digital Library

one-time fee to cover the reasonable costs of the export. However, content that is not paid for within six months of the storage invoice date will be considered abandoned and may be subject to deaccessioning.

Unless specifically requested by the content owner (e.g., in accordance with institutional dataset retention policies), content deaccessioning ultimately occurs at CDL's discretion, as the Library may choose to cover storage costs for a collection even if the content owner is unable to provide adequate funds.

If content has been marked for deaccessioning, CDL will:

- Consult with the content owner to devise an exit strategy from Merritt's cloud storage to another storage solution, be it cloud, on-premises NAS, or device-based storage.
- In the case of a device-based storage transfer, device costs will be the responsibility of the content owner. Egress fees associated with cloud storage, while not expected, will be covered by CDL.
- Consultation will occur over the initial time period of six months, with the option to extend at CDL's discretion.
- Optionally, CDL may choose to cover collection storage costs for up to one year during the consultation process, and/or while a content owner seeks additional funding.
- If additional funding is acquired by the content owner and is grant-based (from a private funding organization), CDL will work with the owner to make use of these new grant funds over an agreed-upon period of time. CDL cannot make use of grant funds that stem from California government or U.S. Federal government grants.

TAKE-DOWN REQUESTS

The procedures for responding to DMCA-compliant take-down requests are defined as part of [CDL's general terms of service](#).

INDEMNIFICATION

CDL makes no representations or warranties with respect to Merritt, and disclaims any liability arising out of their use. Neither CDL nor Merritt users shall be liable for any indirect, special, incidental, punitive or consequential damages arising out of that use. Liability for direct damages is limited to the

California Digital Library

dollar amount of the fee paid for the service. By making use of Merritt, users are indemnifying, defending, and holding harmless CDL, its officers, employees, and agents from and against any liability and damages, including any reasonable attorney's fees, that arise from that use. No limitation of liability set forth elsewhere in these terms applies to this indemnification; further, this indemnification shall survive the termination of these terms.

Part IV: Format Guidelines, Versioning and Persistent Identifiers

FORMAT GUIDELINES

To ensure Merritt preservation strategies can be administered across multiple genres, formats, and packages, CDL provides [Guidelines for Digital Objects](#). Once under secure management, all content is accessible for ongoing review and enrichment by campus-based curators, collection managers, and RDM specialists to maintain and increase its curatorial value and provide a higher level of assurance of its ongoing availability and usability.

CONTENT VERSIONING

Merritt is a strongly versioned repository. Any changes to data or metadata automatically results in the creation of a new version of a digital 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 dataset version. All previous object versions can be retrieved through the Merritt user interface and API.

PERSISTENT IDENTIFIERS

All objects managed in Merritt are assigned unique, persistent Archival Resource Key (ARK) identifiers using CDL's [EZID](#) service. Merritt object landing pages prominently display the object's actionable persistent identifier(s) for use in citations.

Part V: Service Providers, Storage Costs and Availability

SERVICE PROVIDERS

Merritt relies on internal and external service providers for primary and replication storage in its preservation system as well as its compute hosts.

San Diego Supercomputer Center (SDSC)

SDSC provides [Qumulo storage](#), which incorporates an S3-compatible API layer known as MinIO. The Qumulo file storage system 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. Furthermore, SDSC's cloud storage is routinely subject to [Nessus](#) scans, a professional auditing service that probes for vulnerabilities and malware.

For a description of agreements defining the terms of the contractual arrangements between CDL and SDSC, please see the [SDSC Service level Agreement](#).

Amazon Web Services (AWS)

AWS [S3](#) and [Glacier](#) are used for preservation storage, while database hosting is provided through use of [RDS](#), and virtual server hosting via [EC2](#). All of these services are located on the West coast (Oregon).

For a description of agreements defining the terms of the contractual arrangements between CDL and Amazon, please see:

[Amazon AWS customer agreement](#)

[Amazon AWS S3/Glacier service level agreement](#)

[Amazon AWS EC2 service level agreement](#)

AWS complies with a number of regulatory and professional IT standards and certification programs, including CSA, FERPA, FISMA, HIPAA, ISO 9001, 27001, 27017, SOC 1, 2, 3, and others: [AWS Compliance](#).

Wasabi Cloud Storage

[Wasabi Hot Cloud Storage](#) is used as preservation storage for an additional object copy and is located on the East coast (Virginia). The customer agreement that defines the terms of the contractual

California Digital Library

relationship between the University of California Office of the President and Wasabi, and the Wasabi privacy policy are available here:

[Wasabi Technologies Customer Agreement](#)

[Wasabi Privacy Policy](#)

Wasabi complies with a number of regulatory and professional IT standards and certification programs including HIPAA, FERPA, SOC 2, ISO 27001 and PCI-DSS: [Wasabi Compliance](#).

STORAGE COSTS

Merritt operates on a partial cost-recovery basis. There is no service fee for their use, but CDL recoups its costs for provisioning preservation storage, which is typically billed at the campus level. The current nominal pricing is \$150/TB/year, but this is prorated to reflect actual daily storage usage.

Usage accounting is based on the sum total of byte-days of usage over the year, assessed at \$0.000000000000411 per byte-day ($\$150/\text{TB}/\text{year} \div 1,000,000,000,000 \text{ bytes}/\text{TB} \div 365 \text{ days}/\text{year}$). The reliance on byte-day accounting means that contributors do not need to be concerned about the timing of their deposits. 1 TB deposited on the first day of a billing year and saved for the entire year will accrue a cost of \$150 ($1 \text{ TB} * 365 \text{ days} * 1,000,000,000,000 \text{ bytes}/\text{TB} * \$0.000000000000411/\text{byte-day}$). That same 1 TB deposited on the last day of the billing year will cost only \$0.41 ($1 \text{ TB} * 1 \text{ day} * 1,000,000,000,000 \text{ bytes}/\text{TB} * \$0.000000000000411/\text{byte-day}$).

The billing year is aligned with the University of California fiscal year, July through June. Billing for the previous year's storage usage is billed early in the subsequent year, and is payable within 60 days of billing.

Any changes to the Merritt fee structure will be provided to content owners at least 60 days prior to the effective date of the change.

AVAILABILITY

Merritt is available on a nominal 24x7x52 basis. The current status of Merritt availability can be found on the [CDL system status page](#).

California Digital Library

Whenever possible, major service outages for purposes of preventative maintenance and periodic enhancement are scheduled outside of normal business hours, Monday - Friday, 8:00 AM - 5:00 PM PT, and announced two weeks before the scheduled outage. In some cases unanticipated conditions may require immediate intervention without prior announcement in order to prevent damage or loss to managed content. However, Merritt's architecture has been carefully designed for robust fault-tolerance to minimize this necessity. Most diagnostic and maintenance activities can take place without any service interruption.

Part VI: Appendices

New Collection Intake Form

A new [collection intake form](#) is filled out for each new Merritt collection to be established.

Contact

Merritt administrators may be contacted at uc3@ucop.edu, which automatically opens in a new issue in CDL's internal ticketing system.

To report an urgent problem with Merritt, call the [CDL Help Line](#) at (510) 987-0555.