



**Digital
Preservation
Policy Plan**
*Nicolina Gallo in
collaboration with
UC Digital Library
2023*

As my previous paper pointed out, the UC Berkeley Library's Digital Collections (UCBDC) is a massive, noteworthy collection of distinct, signature, rare and exceptional literature, articles, dissertations, photographs, periodicals, audiovisual material and much more. It functions with the sole objective of conserving, safeguarding, and offering access to its collection of digital assets to support the persistent value of academic research, exploration, and advancements. Though I was not able to acquire a copy of the existing digital preservation policy, I was nonetheless able to research multiple other related sources to discover many of the details under which it currently functions. Additionally, I was able to review several online examples of both the Digital Preservation Coalition's Rapid Assessment Model (DPC-RAM) and the Canadian Heritage Information Network (CHIN,) framework assessments. Therefore this paper will present the information I *do* have of the UCBDC's Digital Preservation Policy in an examination of the way in which the institute's foresight in preservation leads policy design, explores the symbiotic relationship between the institution's overarching policies and its mission, evaluates how the existing preservation policy aligns with the institution's strategic vision, while attempting to identify potential obstacles and considerations, and categorize the diverse range of holdings managed by the institution.

Salwa Ismail, the Associate University Librarian for Digital Initiatives and IT in the Digital and Collaborative Services Division, whom I worked with on the earlier paper, was only available for a brief period to go over UCBDC's existing, and potential, digital preservation policies. Thus, I chose to use the CHIN framework as the UCBDC's Digital Repository, also known as, 'Merritt,' is based on the aforementioned metrics. My assessment concluded the following:

Assessment of the UC Berkeley Digital Collection's Digital Preservation Policy

This Digital Preservation Policy outlines the intent, strategies, and responsibilities associated with the Merritt digital preservation repository, which is a service provided by the California Digital Library (CDL). It provides context surrounding the repository, defines CDL and other contributor responsibilities, and outlines the systematic application of technology for preservation assurance. This policy also includes guidelines for content ingest, preservation, stewardship, and access across a multitude of scenarios.

Context and Services

The California Digital Library supports the University of California community in its pursuit of scholarship and public service. The Merritt digital preservation repository is a core CDL service available for managing, preserving, publishing, and sharing valuable digital content created by the university community. Merritt complies with the general terms of service provided by CDL, as well as the terms presented within this policy guide. (Fortnery, 2020)

Digital Preservation Strategy

Digital preservation encompasses a human element, a strategic element, actors, policies, procedures, and technologies aimed at ensuring access to digital content over time while maintaining its usability, authenticity, and integrity. The Merritt repository focuses on providing bit-level preservation, safeguarding the integrity of each file stored by preserving the bits that encode the digital materials. To strengthen its preservation strategy, Merritt maintains three copies of all files and digital objects in the system through external storage providers. The content is stored across multiple cloud storage providers in different geographic regions, minimizing the risk of data loss due to disaster events. Merritt follows the Open Archive Information System (OAIS) reference model and actively manages the preservation process, including automated classification of file types, fixity checking, and daily system consistency checks. (Lopatin, 2022, para. 4-8)

Privacy, Accessibility, and Responsibilities

Merritt complies with CDL's privacy policy, respecting and protecting the privacy of all users in accordance with federal and state laws and University of California policies. (CDL, 2020, para. 1-

Accessibility

Merritt adheres to CDL's accessibility policy, ensuring that its IT environment is accessible to a broad population, allowing individuals to access and benefit from the University's electronic programs and services. (Lopatin, 2022, para. 11)

Contributor Responsibilities

Contributors to Merritt acknowledge that they have followed applicable laws, regulations, policies, ethical concerns, and disciplinary best practices when creating or acquiring content. They grant CDL the right to preserve, copy, enhance, federate, create derivatives, and provide access to contributed content, subject to access controls defined by curators. Contributors exhibiting inappropriate behavior may face user privilege restrictions.

Merritt is not suitable for managing content containing clinical or personally identifiable information (PII) as that would denote a HIPAA/HITECH, FERPA, or similar violation. Content with PII must be redacted or anonymized before submission to Merritt. (Lopatin, 2022, para. 12)

CDL Responsibilities

CDL accepts, manages, and provides access to digital content to support the University's mission. CDL ensures the highest level of preservation assurance based on content characteristics, metadata availability, and organizational priorities. CDL complies with its privacy policy and is committed to securing bit-level preservation for all content. Additionally, CDL offers consultation and guidance on acquiring or creating digital content in a manner compatible with future preservation needs. (Lopatin, 2022, para. 14)

Format Guidelines, Versioning, and Persistent Identifiers

Format Guidelines

CDL provides guidelines for digital objects to ensure preservation strategies can be applied across various genres, formats, and packages. Content is accessible for ongoing review and enrichment by

curators and collection managers to maintain its curatorial value and ensure ongoing availability and usability. (CDL, 2021, para. 24)

Content Versioning

Merritt is a strongly versioned repository, creating frequent new varieties of digital objects when changes occur. The versioning relies on minimum storage duplication. Previous versions of objects can be retrieved through the Merritt user interface and API. (Lopatin, 2022, para. 23)

Persistent Identifiers

All objects managed in Merritt are assigned unique, persistent Archival Resource Key (ARK) identifiers using CDL's EZID service. Persistent identifiers are prominently displayed on Merritt's object landing pages for citation purposes. (Lopatin, 2022, para. 25)

Service Providers, Storage Costs, and Availability

Service Providers

Merritt relies on internal and external service providers for primary and replication storage, as well as compute hosts. The San Diego Supercomputer Center (SDSC) provides storage, while Amazon Web Services (AWS) and Wasabi Cloud Storage are used for preservation storage. (Lopatin, 2022, para. 26)

Storage Costs

Merritt operates on a partial cost-recovery basis, and storage costs are prorated based on actual daily storage usage. The pricing is currently set at \$150 per terabyte per year, but the actual cost is calculated using a byte-day accounting system. Billing occurs annually, aligning with the University of California fiscal year. (Lopatin, 2022, para. 27)

Availability

Merritt is available 24 hours x7 days a week and 52 weeks per year, with scheduled maintenance and enhancements typically conducted outside normal business hours. Major service outages are announced in advance whenever possible, and Merritt's architecture is designed for fault-tolerance to minimize service interruptions. (Lopatin, 2022, para. 39)

The Digital Preservation Policy for the Merritt Digital Repository outlines the repository's intent, strategies, and responsibilities for preserving digital content. It aligns with the UC Berkeley Library's mission to support the University of California community in its research, teaching, and public service activities, ensuring the long-term accessibility, usability, and integrity of the digital resources managed by Merritt. While exploring the interplay between Merritt's digital preservation preparedness and the development of its policies, the prescience of continuity in the preservation is clear in the way that it orchestrates the policy's design. The capabilities, infrastructure, and practices of the repository give rise the scheme of the guidelines themselves. Factors such as its preservation strategy, technological infrastructure, compliance with standards, and expertise in digital preservation make up the foundation of Merritt's preservation. The policies are designed to address the specific needs and requirements of the digital preservation activities conducted within the repository. For example, the preservation strategy of bit-level preservation and the use of

multiple storage providers are reflected in the policy to ensure the long-term integrity and availability of digital content. Additionally, Merritt's adherence to standards, such as the Open Archive Information System (OAIS) reference model, influences the policy to ensure compliance with recognized best practices and the interoperability necessary for other digital preservation systems. The organizational structure, responsibilities, and resources available within Merritt demonstrate clear strategies and expectations for employees and contributors involved in the preservation process.

The mission of Merritt is to preserve, protect, and provide access to the digital collections, archival records, and other materials it holds to sustain the ongoing integrity and value of intellectual inquiry, scholarship, and achievement. There is an unambiguous relationship between Merritt's wider policies and its mission. The overarching purpose and goal of the repository is to secure and make accessible the UC Berkeley-stored wealth of information and resources. digital collections and archival records. The policies, thus serve as the guidelines and strategies for achieving that mission. The policies maintain safeguards for authenticity, quality, and usability. Over time, they outline specific actions, procedures, and technologies that fulfill those preservation responsibilities. The repository currently has multiple relevant policies specific to its mission. However, I've outlined a few of the more notable ones below:

- 1.) **Digital Preservation Strategy:** This policy outlines the systematic application of technology and procedures to ensure the long-term continuity of the digital materials. It includes the use of bit-level preservation, mandates that multiple file copies be preserved in a separate location, calls for fixity checking, and adherence to the Open Archive Information System (OAIS) reference model. (Lopatine, 2022, para. 9)
- 2.) **Privacy, Accessibility, and Responsibilities:** This policy articulates the privacy and accessibility aspects of operations. It requires compliance with all privacy laws and regulations, ensures the accessibility of digital content, and allots administration of any associated fines toward both Merritt contributors *and* the California Digital Library (CDL, 2020, para. 2-9).
- 3.) **Contributor Responsibilities:** This guideline lays out the responsibilities of Merritt contributors. Emphasizing deference to the appropriate ethical, legal, and regulatory matters, it also stresses the best practices related to content creation and acquisition and includes provisions for intellectual property rights and scholarly discourse. (Lopatine, 2022, para. 13)
- 4.) **Format Guidelines, Versioning, and Persistent Identifiers:** This policy directs the route of digital objects in Merritt, and includes suggestions for format selection, content versioning practices, and the assignment of persistent identifiers (ARKs) to ensure reliable identification and citation of digital objects. (CDL, 2021, p. 6-7)

These and many further policies within what I've been able to see of the existing Digital Preservation Policy that unanimously addressed the various elements of digital preservation, privacy, accessibility, content management, and responsible stewardship, all of which are essential to fulfilling Merritt's mission.

Many of the policies interrelate to one another with fluidity. Several highlight the comprehensive strategy and approach for preserving digital content in Merritt. They emphasize the

importance of bit-level preservation, multiple copies of files, fixity checking, and adherence to the ISO 14721 OAIS reference model. These strategies align with Merritt's mission to preserve and protect digital collections, ensuring their enduring value for intellectual inquiry and research. The principles surrounding privacy and accessibility comply with applicable laws and regulations, while also advancing an accessible IT environment that allows a broad population access to the digital collections for ongoing intellectual discovery. (Meltzer, 2011) CDL commits to preserving content at the bit level, offering expert consultation, securing a high level of preservation assurance within resources available while the policies of formatting and versioning, and the assignment of persistent identifiers establish effective organization and management. Both go further in reinforcing Merritt's mission of preserving the digital collections while providing enduring value for academia. From the portion of I was able to study, the existing preservation policy does indeed reflect the “grand scheme” of the overall institution. The context and services section sets the tone for the subsequent guidelines and establishes the broader framework for Merritt's operations. The digital preservation strategy outlines the procedures employed by Merritt, while highlighting the importance of born-digital content over time and the equally essential goals of authenticity and integrity. The policy stresses the use of available -level preservation, fixity checking, and adherence to the ISO 14721 OAIS reference model. (Fortney, 2020, para. 1-3) The policy concerning service providers, storage costs, and availability conveys specific information on the infrastructure and service providers utilized by Merritt for primary and replication storage. It contains details about the various systems, employed, such as Qumulo storage, provided by the San Diego Supercomputer Center, AWS S3 and S3 Glacier storage, and Wasabi Hot Cloud Storage. (Lopatin, 2022, para. 27-32)

Given that I only have access to a certain part of the policy, I can't say with any certainty whether the following is completely accurate, however, despite Merritt's existing preservation policy maintaining a solid foundation for digital preservation and repository management, there are a few potential issues that could be considered:

- There is a lack of explicit mention of long-term sustainability: It could be helpful to include information about the institution's commitment to long-term funding and resource allocation.
- It may also be beneficial to have transparent guidelines that ensure the inclusion of valuable and relevant digital materials while considering factors like scholarly impact, research significance, and collection priorities.
- While the policy makes mention of specific format guidelines, there is no detailed recommendation with which to approach format obsolescence and migration
- The policy mentions distributing file copies across multiple cloud providers to mitigate risk, but there is little discussion on disaster recovery or any mention of off-site storage.
- There is a brief section that touches on offering consultation and guidance, but there is limited emphasis on community engagement and collaboration.

Based on the information available, it was nearly impossible to uncover any information about Merritt's holdings. Yet it maintains a collection that spans a diverse array of disciplines and levels of professionalism, education, creativity, and beyond. The assemblage of materials involved, and

topics covered is nearly infinite. Composed of digitized materials, born-digital content, research data, scholarly publications, institutional records, and other digital assets, Merritt also appears to house a large collection of archival records, research data and cultural ephemera. The way in which the very nature of preservation is treated and conceived of here at this incredible institution is a direct proponent of its policy creation. With the wealth of information contained within the Digital Collections, even the most robust preservation would likely fall short of every need. Yet the ongoing upkeep, expansion, services, and increasingly more relevant technology that continues to be integrated into the Digital Preservation Policy, I feel this is a very extensive, well-thought out, easily upgraded set of guidelines for an equally poignant and diverse array of materials.

Bibliography:

Academic Preservation Trust, Chronopolis, et. al. (May 2022). *Digital Preservation Declaration of Shared Values*. https://dpscollaborative.org/shared-values_en_v3.html

California Digital Library (CDL). (07, Aug., 2018). *CoreTrustSeal requirements 2017-2019*. <https://www.coretrustseal.org/wp-content/uploads/2018/08/UC3-Merritt.pdf>

California Digital Library (CDL). (n.d) *Merritt: A trusted, cost-effective digital preservation repository*. <https://merritt.cdlib.org/>

California Digital Library (CDL). (n.d). *Merritt Preservation System*. <https://www.ucop.edu/information-technology-services/files/uc-tech-awards/2022/cdl-merritt-preservation-system.pdf>

California Digital Library (CDL) Merritt Preservation System. (n.d) *Operational excellence award submission*. <https://www.ucop.edu/information-technology-services/files/uc-tech-awards/2022/cdl-merritt-preservation-system.pdf>

California Digital Library (CDL). (13. Aug., 2019) *Mission, vision, and values*. <https://cdlib.org/about/mission-vision-and-values/>

California Digital Library (CDL). (17, Dec., 2020). *Privacy Policy*. <https://cdlib.org/about/policies-and-guidelines/privacy-policy/>

California Digital Library (CDL). (June, 2021). *CDL guidelines for digital objects*. <https://cdlib.org/wp-content/uploads/2019/01/GDO.pdf>

Canada. Department of Canadian Heritage. (08, Jun., 2021). *Digital Preservation Plan Framework for Cultural Heritage Institutions*. <https://www.canada.ca/en/heritage-information-network/services/digital-preservation/plan-framework-museums.html>

Consultative Committee for Space Data Systems. (2002). *Reference model for an open archival information system (oais)*. CCSDS Secretariat. <http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>.

Duranti L. Preston R. InterPARES Project & Associazione Nazionale Archivistica Italiana. (2008). *International research on permanent authentic records in electronic systems (inter pares) 2: Experiential interactive and dynamic records*. CLEUP.

Fortney, K.(22., Sep., 2020). *California Digital Library's terms of service*. <https://cdlib.org/about/policies-and-guidelines/terms-conditions/>

Lavoie B. F. Digital Preservation Coalition & Charles Beagrie Ltd. (2014). *The open archival information system (oais) reference model: Introductory guide* (2nd ed.). Digital Preservation Coalition. <https://doi.org/10.7207/twr14-02>

Lopatin, E. (16, Sep., 2020). Merritt. <https://cdlib.org/services/uc3/merritt/>

Lopatin, E. (30, Sep., 2022). Merritt frequently asked questions. <https://cdlib.org/services/uc3/merritt/merritt-frequently-asked-questions-faq/>

Meltzer, E. (26, May, 2011). *Websites: Make them accessible to everyone*. <https://cdlib.org/cdinfo/2011/05/26/websites-make-them-accessible-to-everyone/>

OAC, Calisphere, (24, Aug., 2021). *UC shared DAMS (Nuxeo) and ArchivesSpace*. <https://cdlib.org/services/pad/ucdams/>

RLG/OCLC Working Group on Digital Archive Attributes Research Libraries Group & OCLC. (2002). *Trusted digital repositories : attributes and responsibilities: An rlg-oclc report*. RLG.

Schaefer, S., Chodacki, J., Ismail, S. et. al. (10, Aug., 2020). *University of California digital preservation strategy working group: Phase two report*. <https://escholarship.org/uc/item/3kq6h67m-main>

Smith, E., Chodacki, J. Elings, M, et. al. (10, Apr., 2019) *UC digital preservation strategy working group: Phase one report*. <https://escholarship.org/content/qt80v318pm/qt80v318pm.pdf?t=qgr1nv>

UC3. (2022). *Merritt digital preservation repository policies and user guidelines version 2.0*. <https://cdlib.org/wp-content/uploads/2022/09/Merritt-Preservation-Repository-Policies-User-Guidelines-v2.pdf>

UC Berkeley Library. (2022). *UC Berkeley library organization chart*. <https://www.lib.berkeley.edu/sites/default/files/library-organization-chart.pdf>

UC Berkeley Library. (2023). *About digital collections*. <https://www.lib.berkeley.edu/find/digital-collections/about>

WAG. (2011). *UC Berkeley Library website mission, goals & scope*. <https://www.lib.berkeley.edu/Staff/lit/mission>

Yeung, T. and University of Calgary. (2004). *Digital Preservation: Best Practice for Museums*. Minister of Public Works and Government Services, Canada.

[https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/68062/HortonR_Operationalizing
DAFDigitalArchives_2022.pdf?sequence=1](https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/68062/HortonR_Operationalizing_DAFDigitalArchives_2022.pdf?sequence=1)