

Project director	David Moles
Institution	University of California, Berkeley
Project funding	Outright funds: \$50,000
Application information	Project site location: CA-12

One Link in an Unbroken Chain of Memory

Modernizing Digital Preservation at UC Berkeley

Synopsis The University of California, Berkeley, is requesting a Foundations award under the Humanities Collections and Reference Resources program (HCRR) to support planning, assessment, and pilot activities for developing a digital preservation policy, for establishing comprehensive intellectual control of digital content previously deposited by the UC Berkeley Library for preservation in the California Digital Library's Merritt repository, and for exploring technical and organizational options for the continued long-term preservation of that content, such as Merritt modernization and/or migration of UC Berkeley content to an alternative digital repository platform.

Requested funding \$50,000.00

- \$35,714.29 direct salary costs (3 mos. FTE)
- \$14,285.71 indirect costs (per UC Berkeley negotiated on-campus "other sponsored activities" rate of 40%)

Narrative

There is no end for digital preservation. The best one can hope for is to be one link in an unbroken chain of memory.

—Owens, *The Theory and Craft of Digital Preservation*¹

Significance

The University of California, Berkeley (hereafter, UC Berkeley or UCB) is a public, coeducational research university, offering a variety of programs from bachelors' degrees through research and professional doctorates. Founded in 1868, it is the oldest and, by number of students, the largest² of the University of California's ten campuses.³ The UC Berkeley Library, founded in 1869, comprises 22 campus libraries holding more than 13 million volumes.⁴ Humanities materials include not only books and manuscripts but also Egyptian papyri, 20th-century ephemera, artworks and vintage costumes. The Bancroft Library, UC Berkeley's primary special collections library, holds one of the country's largest and most heavily used collections of rare books, manuscripts, and unique materials.⁵

In the digital realm, the UC Berkeley Library has already digitized nearly 2.5 million items, and UC Berkeley is by far the largest contributor to the Merritt digital preservation repository at the University of California's California Digital Library (CDL), with Berkeley materials accounting for more than 2 million of Merritt's roughly 3 million digital objects.⁶ This achievement, however, is only the beginning of progress toward the Library's long-term goal: to digitize all two hundred million items in its special collections, and make them accessible to the public.⁷

While the UC Berkeley Library's current digital preservation infrastructure, organization, and processes have served it well to this point, if preservation is to keep pace with the Library's ambitious digitization efforts, it is time to incorporate fifteen years of digital preservation lessons learned both at UC Berkeley and at CDL, as well as the best practices developed by a digital preservation community now considerably more mature than when Berkeley's pioneering efforts began. With this project, the UC Berkeley Library will initiate an assessment and planning process with the intertwined goals of formalizing digital preservation policy at Berkeley, taking stock of digital collections currently being preserved, and exploring future technological and organizational directions for preservation—readying the tools and raw materials needed for the Library to forge the next link in Owens' "unbroken chain of memory."

History, scope, and duration

UC Berkeley lacks a comprehensive digital preservation policy. Historically, digital preservation at Berkeley has been considered a primarily technical problem, with the details of its implementation relegated to the Library IT department, and with minimal oversight from library stakeholders. Library IT, meanwhile, has treated Merritt as an adjunct to the Library's other digital platforms, such as the patron-facing Digital Collections website⁸ and its underlying TIND DA digital asset management system, rather than as a patron-facing access platform in its own right.⁹ Should a patron locate an image on the Digital Collections website and wish to acquire a preservation-quality version, the patron must contact the Library to request it.¹⁰ A Library IT staff member must then examine the metadata for the public-facing object, determine what Merritt collection is likely to contain the preservation copy, and locate that copy in Merritt.¹¹ At the same time, the choice of items to digitize and preserve has been largely ad-hoc, with priorities often driven by funding windfalls or by short-term digitization needs.¹²

In 2018-2019, the UC Berkeley Library initiated its Digital Lifecycle Program (DLP) as a strategic project with Kathryn Stine as primary advisor, culminating in the selection and launch of the TIND DA digital asset management system in fall 2019.¹³ Hiring of a permanent, full-time DLP program manager was postponed due to the COVID-19 pandemic, and has been further postponed due to the impending retirement of

University Librarian Jeff Mackie-Mason. Nonetheless, this project, led by Ms. Stine as project manager, will build on the work done in the initial DLP effort, and is in line with the future priorities and strategic directions that project identified, of coordinating all aspects of digitization, digital asset management, and digital preservation and scaling up to support a throughput of 3 million items per year.¹⁴

Methodology and standards

Standards, research, and precedents

Development of CDL's Merritt repository, which serves as the UC Berkeley Library's primary digital preservation platform, began in 2010, and was roughly contemporaneous with the founding of the National Digital Stewardship Alliance (NDSA).¹⁵ Though Merritt's original design was informed by the earliest versions of the Open Archival Information System (OAIS), and though both CDL and UC Berkeley are members of the NDSA,¹⁶ neither the Merritt software platform nor the processes surrounding its usage at Berkeley have kept pace during the intervening years with the evolution of standards and best practices in the digital preservation community. While this project will incorporate current process where possible, and while further investment in Merritt may well be the technical direction the project eventually recommends, the planning process will begin with a clean slate, following the policy development guidelines formulated by the Digital Preservation Coalition and evaluating practices against the most recent set of criteria published by the NDSA.¹⁷

Required skills & team selection

Familiarity both with contemporary digital preservation guidelines and practice, and with the current UC Berkeley / CDL digital preservation infrastructure, is therefore critical to this project. Additionally, evaluation of digital preservation technical solutions requires expertise both in library systems and in enterprise information technology. David Moles, project director and technology strategist, is former technical lead for Merritt and former Head of Applications for UC Berkeley Library IT, an experienced technology manager with considerable experience in software evaluation and selection, both in the cultural heritage domain and outside it. In addition to working with digital preservation systems at CDL and at Berkeley, Mx. Moles has also studied digital preservation at Rutgers and at the University of Wisconsin, and has good relationships with digital preservation experts at the Digital Preservation Coalition, at the National Archives, and at Harvard.

As noted by Trevor Owens, "digital preservation is people."¹⁸ Successful execution of this project therefore requires not only a thorough understanding of the technicalities of digital preservation and the history and practice of digital preservation at Berkeley,

but also the ability to work with, elicit information from, and build consensus among the diverse individuals and populations involved in that history and practice, from Berkeley librarians and archivists and Library IT personnel, to CDL technical and service management staff; familiarity with the politics and personalities involved is also of considerable value. Kathryn Stine, project manager and metadata strategist, has worked both for CDL and the UC Berkeley Library and is highly respected at both, while having a track record of delivering complex projects requiring coordination across multiple constituencies.

Jackie Gosselar, advisor, brings extensive knowledge of the Berkeley and larger UC library organizations and a unique perspective on large-scale system implementation and migration, thanks to their work on UC's Systemwide Integrated Library System; Mx. Gosselar is also a respected librarian and manager with a background in cataloging and metadata. Mx. Gosselar can therefore provide invaluable feedback on project proposals and preliminary results before they are circulated to the broader stakeholder community.

Information gathering

Information gathering will consist primarily of personal interviews with relevant staff, librarians, and archivists at Berkeley and at CDL, of technical analysis of digital content currently being preserved, and of consultation with outside experts at peer institutions.

Collaboration and decision-making

The core project team, consisting of Mx. Moles and Ms. Stine, will meet regularly both face to face and via teleconference, as appropriate. Decisions will be collaborative. In particular, recommendations will be circulated to stakeholders, with the goal of ensuring positive-sum outcomes for all concerned; in the event of disagreements, Ms. Stine and Mx. Moles will work with stakeholders to achieve consensus, in accordance with UC Berkeley Library culture. In the event consensus cannot be reached, however, final decision authority will rest with the University Librarian.¹⁹

Project deliverables

Deliverables for this project include:

- a comprehensive statement of Berkeley's digital preservation policy, in draft form suitable for high-level stakeholder review and signoff

- an intellectual control strategy for Berkeley's preserved digital content, incorporating standards for descriptive, technical, and preservation metadata as well as internal and external identifier assignment, digital collection scope and structure, and digital object structure, and a project plan for bringing existing content into line with this strategy
- identified technical and cost requirements for Berkeley's next-generation digital preservation system, and a matrix of candidate platforms / products to assess against these requirements

Development of this comprehensive digital preservation policy will raise awareness of digital preservation issues in the UC Berkeley Library community, while its adoption will provide a stable framework for future digital preservation decisions. The implementation of a purposeful intellectual control strategy, and the evaluation and eventual selection and deployment of a next-generation preservation platform, will similarly provide a stable environment for future digital preservation actions.

Sustainability of project outcomes and digital content

UC Berkeley and the UC Berkeley Library have been in existence for more than 150 years. For more than a decade of that time, digitization and digital preservation have been among the Library's core strategic efforts. The Library's current digital preservation system, the California Digital Library's Merritt, is a CoreTrustSeal certified trustworthy data repository,²⁰ and no system, workflow, or organization will be considered by this project that cannot also qualify for that certification.

Outreach

The goal of this project is to facilitate the expansion and long-term technological and financial sustainability of the UC Berkeley Library's already highly successful digital collections efforts, in as transparent a manner as possible—transparent in the sense that the project process and outputs will be public, and transparent in the sense that any system migrations will be seamless and any metadata alterations additive, so as not to disrupt ongoing use of the collections. Public evidence of project success will reside in the new digital projects facilitated by this work, and in making access to preserved digital collections simpler and more reliable.

Work plan

The project will address three areas of concern: developing a digital preservation policy, establishing comprehensive intellectual control of digital content already under

preservation, and exploring future technical and organizational directions for the next stage of digital preservation at Berkeley. Work will be done primarily by project manager and metadata strategist Kathryn Stine and by project director and technology strategist David Moles, each of whom will dedicate roughly 25% of their time to this project over an estimated six months from June through November 2025.

June 2025

- **DP policy:** Ms. Stine will identify relevant digital preservation stakeholders and clarify the policy's target audience, draft an initial statement of digital preservation principles and of the scope and objectives of the policy and circulate this statement for review. In parallel, Mx. Moles will document current digital preservation policies and practices and identify policy gaps. Both project leads will work to promote and socialize the policy planning process to ensure stakeholder buy-in.

July-August 2025

- **DP policy:** Ms. Stine will finalize the statement of principles, policy scope, and policy objectives, and begin working with stakeholders to document the organizational context of digital preservation at Berkeley, and to identify areas of contention or contrary assumptions about priorities, principles, or responsibilities. Mx. Moles will work with Library IT to delineate boundaries between digital preservation and related activities such as digital asset management, storage management, and disaster recovery.
- **Control:** Mx. Moles, working with Library IT and with the Merritt team at the California Digital Library, will begin developing an inventory of digital content already in Merritt and gathering data to assess the quality of metadata, identifier assignment, and physical / intellectual arrangement.
- **Future:** Mx. Moles will begin working with stakeholders to identify use cases for control over and access to preserved digital content, including "dark" collections.

September-October 2025

- **DP policy:** Ms. Stine will work with stakeholders to resolve areas of disagreement and begin compiling a draft digital preservation policy.
- **Control:** Ms. Stine and Mx. Moles will draft target standards for digital object metadata, internal and external identifier assignment, collection structure and object structure, and circulate these standards for review.
- **Future:** Mx. Moles will consult with stakeholders, with Library IT, and with colleagues at peer institutions to determine technical desiderata and expenditure limits for Berkeley's next-generation preservation platform and survey the preservation technology and economics landscapes.

November 2025

- **DP policy:** Ms. Stine will complete the draft policy and circulate it for review.
- **Control:** Incorporating feedback on the draft target standards, Ms. Stine and Mx. Moles will develop a preliminary project plan for enhancing metadata and identifier assignment and improving arrangement of existing content.
- **Future:** Mx. Moles will complete the landscape survey and produce a matrix of potential next-generation preservation platforms with associated cost estimates.

Project personnel and advisors

David Moles, project director and technology strategist, is the former Head of Applications for the UC Berkeley Library, and the former technical lead for the Merritt team at the California Digital Library's UC Curation Center. They have an M. Phil. in history from the University of Oxford, a BA in language studies from the University of California, Santa Cruz, and are currently completing a Master of Information degree in archives and preservation at Rutgers. Mx. Moles is an experienced technical project lead whose work has included numerous data migration, platform migration, and metadata enhancement and conversion projects.

Kathryn Stine, project manager and metadata strategist, is the UC Berkeley Library's current Digitization Project Planner, and the former Senior Product Manager for Digitization and Digital Content at the California Digital Library. She has both an MSI in archives and records management / museum studies and an MFA in printmaking and mixed media from the University of Michigan, and a BA in art practice from UC Berkeley. In her career as a digital library strategist Ms. Stine has been responsible for multiple innovative and collaborative digital collections systems, services, policies, and programs, including developing the UC Berkeley Library's Digital Lifecycle Program plan, and has considerable experience in metadata analysis and metadata strategy.

Jackie Gosselar, advisor, is the Systems and Discovery Services Librarian for the UC Berkeley Library. Mx. Gosselar has both an MA in library and information science and a BA in anthropology / art history from the University of Wisconsin, Madison. Mx. Gosselar manages the UC Berkeley Library's Systems and Discovery Services team, and has extensive knowledge of both patron- and staff-facing user experience issues, as well as considerable experience with system migration and metadata reconciliation; they have also worked as a cataloger and metadata analyst.

Notes

- ¹ p. 200.
- ² Excluding medical school students, as UC Berkeley lacks a medical school, that role being filled by nearby UC San Francisco. If the medical school is included, UCLA's student body is slightly larger.
- ³ UC Berkeley (n.d.); UC (2022).
- ⁴ UC Berkeley Library (2021, 2022).
- ⁵ UC Berkeley Library, "Bancroft Library" (n.d.).
- ⁶ UC Berkeley Library (2022); Moles (2018).
- ⁷ UC Berkeley Library Communications (2020).
- ⁸ UC Berkeley Library, "Digital Collections" (n.d.).
- ⁹ Until recently, the structure of digital objects in Merritt did not even reflect the structure of digital objects in the DAMS; individual scanned pages of the same book, for instance, while managed together as a single object in the DAMS, would be stored in Merritt as separate objects.
- ¹⁰ A process neither standardized nor documented. In fact, there is nothing on the Digital Collections website to even make patrons aware that a preservation-quality image may exist.
- ¹¹ Traditionally, the Berkeley approach to making it possible to retrieve objects preserved in Merritt has been to derive the filename for a scanned image from the catalog ID and call number of the physical item being scanned, and embed that filename in the title of the digital object; metadata for individual objects is almost nonexistent, even by Merritt's minimal, Dublin Core-based "who / what / when / where?" standards. (For details of Merritt's approach to metadata, see Lopatin (2022).)
- ¹² Examples include the digitization of pre-1912 Chinese materials, funded by the Alibaba Foundation (Haugan (2021)), and the digitization of course materials for controlled digital lending that began during the COVID-19 pandemic (UC Berkeley Library Communications (2018)).
- ¹³ UC Berkeley Library (2019).
- ¹⁴ UC Berkeley Library (2018, 2019).
- ¹⁵ CDL (2010); NDSA, "About" (n.d.).
- ¹⁶ *ISO 14721:2003* (2003); CoreTrustSeal (2023); NDSA, "Members" (n.d.).
- ¹⁷ DPC (n.d.); NDSA (2019).
- ¹⁸ Owens (2018), p. 9.
- ¹⁹ The position of University Librarian is currently vacant, but is expected to be filled before the commencement of this project.
- ²⁰ CoreTrustSeal (2023).

References

- California Digital Library. (2010). *Merritt in CDL Repository Development Project*.
https://www.digitalpreservation.gov/news/2010/20100527news_article_CDL_merritt.html
- California Digital Library. (n.d.). *UC3 Merritt Home*. <https://merritt.cdlib.org/>
- CoreTrustSeal. (2023). *UC3 Merritt Curated Collection—California Digital Repository*.
https://github.com/CDLUC3/mrt-doc/blob/9b474dd/doc/2023-03-27-California-Digital-Library-UC3-Merritt-Curated-Collection---California-Digital-Library_final.pdf
- Digital Preservation Coalition. (n.d.). *Digital Preservation Policy Toolkit*.
<https://www.dpconline.org/digipres/policy-toolkit>
- Haugan, T. (2021). ‘A long journey’: Richly detailed, fully searchable Chinese treasures will be made available for free online. UC Berkeley Library News & Updates.
<https://www.lib.berkeley.edu/about/news/chinese-treasures>
- ISO 14721:2003—Space data and information transfer systems—Open archival information system—Reference model. (2003). ISO.
<https://www.iso.org/standard/24683.html>
- Lopatin, E. (2022). *Metadata*. <https://github.com/CDLUC3/mrt-doc/wiki/Metadata>
- Moles, D. (2018). *Merritt Collection Management & Replication: Preparatory notes*.
<https://github.com/CDLUC3/mrt-doc/blob/9b474dd/meetings/2018-09-17-merritt-collection-mgmt-replication-prep.md>
- National Digital Stewardship Alliance. (2019). *Levels of Digital Preservation Matrix v2.0*. <https://osf.io/2mkwx>
- National Digital Stewardship Alliance. (n.d.). *About the NDSA*. National Digital Stewardship Alliance - Digital Library Federation. Retrieved March 13, 2024, from <http://ndsa.org/about/>
- Owens, T. (2018). *The Theory and Craft of Digital Preservation*. Johns Hopkins University Press. <https://doi.org/10.1353/book.62324>
- UC Berkeley Library Communications. (2018). *UC Berkeley Library takes key step to expand digital access to its collections*. UC Berkeley Library News & Updates.
<https://www.lib.berkeley.edu/about/news/controlled-digital-lending>

- UC Berkeley Library Communications. (2020). *UC Berkeley Library makes it easier to digitize collections responsibly with novel workflows and bold policy | UC Berkeley Library*. UC Berkeley Library News & Updates. from <https://www.lib.berkeley.edu/about/news/responsible-access>
- UC Berkeley Library. (2018). *Digital Lifecycle Program Launch: Project Charter*. <https://stories.lib.berkeley.edu/wp-content/uploads/2019/04/DLP-Strategic-project-charter-FY19.pdf>
- UC Berkeley Library. (2019). *Digital Lifecycle Program Launch: Status Report*. <https://stories.lib.berkeley.edu/wp-content/uploads/2019/08/DLP-Project-Status-2018-2019.pdf>
- UC Berkeley Library. (2021). *Annual report 2020-21*. [https://www.lib.berkeley.edu/sites/default/files/uc berkeley library annual report 2020-21.pdf](https://www.lib.berkeley.edu/sites/default/files/uc%20berkeley%20library%20annual%20report%2020-21.pdf)
- UC Berkeley Library. (2022). *Library facts and figures*. <https://www.lib.berkeley.edu/sites/default/files/UCBLibraryFacts.pdf>
- UC Berkeley Library. (n.d.). *Bancroft Library*. <https://www.lib.berkeley.edu/visit/bancroft>
- UC Berkeley Library. (n.d.). *Digital Collections*. <https://www.lib.berkeley.edu/find/digital-collections>
- UC Berkeley. (n.d.). *History & discoveries*. <https://www.berkeley.edu/about/history-discoveries/>
- University of California. (2022). *Fall enrollment at a glance*. <https://www.universityofcalifornia.edu/about-us/information-center/fall-enrollment-glance>