

Digital Gap Assessment
Nicolina Gallo
collaboration with
UC Digital Library, 2023

UC Berkeley Library's Digital Collections is an endeavor to advance impassioned access to Berkeley's Library-digitized or acquired digital materials affiliated with digitization and development. Hoping to sustain, secure, and supply opportunities to explore the knowledge within UC Berkeley's respective libraries, the Digital Collections reflect this sentiment further, stating that their aim "encourages and supports research, teaching, and scholarship to advance global knowledge and understanding." (Berkeley Library, 2023, *About*) The goal of this scholarly giant is to ensure ongoing access to both its long-held, and newly-acquired materials, for everyone from its students, professors, and staff to visiting scholars and other researchers, to simple community members throughout the world. In an effort to support the mission of the overarching UC Berkeley library to "help current and future users find, evaluate, use, and create knowledge to better the world," (Berkeley Library, 2023, *Facts & Figures*) the Digital Collections, similarly is a leader and partner in creating and implementing ideas and services that enhance research and scholarship across the world. The library itself seeks to integrate materials it considers "of enduring value for intellectual inquiry, research, and discovery," (UC Berkeley, 2023, *About*) into the historical digital record with an insightful reach while simultaneously observing statutory, contractual, moral, and security-related factors - components it considers critical to improving access to extensive digitized materials and other procured digital content. The Digital Collections is an effort to ensure that UC Berkeley's library materials are "effectively usable today and preserved for future generations." (MacKie & Samberg, 2020)

Similar to the other institutions in the Berkeley library system, the Digital Collections maintain a strong online presence. In addition to their own individual [Digital Collections Home](#) website, they are also affiliated with the extensive presence of the UC Berkeley main library on social media in [Facebook](#), [Instagram](#), [YouTube](#), [Twitter](#), and [LinkedIn](#). Additionally, the opportunity is available online to [Browse the Digital Collections](#), through the online [Digital Collections Portal](#). News and financial information surrounding the Digital Collections comes up now again, though its visibility is far out surpassed by the mentioned of the more ubiquitous primary UC Berkeley Library. During the pandemic, the e-reserves program was created by the Digital Initiative and Information Technology team, the same currently at the base of operations for the Digital Collections. Though respective from the Digital Collections, the e-reserves program was nonetheless an ambitious enterprise, creating free digital versions of textbooks for approximately 39,895 students across 77 academic departments. (MacKie-Mason, 2021, p. 7)

I had the opportunity to interview Salwa Ismail, the Associate University Librarian for Digital Initiatives and IT in the Digital and Collaborative Services Division. She is currently in a supervisory role in digital services, where she specializes in digitization, digital scholarship, digital preservation, digital publishing, research data services, and several various other programs. Her research focuses on interface design and infrastructure with a user-centric and service-oriented approach. Salwa turned out to be an incredibly unique leader in Digital Innovation. She has expertise in applying computational methodologies and agent-based modeling in complex adaptive systems. Within the UC Berkeley Digital Initiative, she oversees the Library's Digital Lifecycle Program, one that increases accessibility to online scholarly material. Salwa personally supervised the pandemic-related operation of digitizing the thousands of materials related to the aforementioned e-reserves program. She was also at the helm when, upon the return of students to campus, the university chose to enhance the mass assemblage of textbooks within the Digital Collections. In a move toward what is likely the future of academic literature, the UC Berkeley Library launched its own "controlled digital lending platform," (Haugan, 2021, para. 29) through which scholars, students, and instructors can now connect to any and all necessary materials. Through what has been labelled 'UC BEARS,' an acronym for the 'UC Berkeley Electronic and Accessible Reserves System,' the open-source platform has emerged from the work of Salwa and her in-house team. The fall semester of 2021 saw UC Berkeley becoming the forerunner in the UC system to introduce a first-of-its-kind system that offered complete electronic versions of textbooks for course reserves. To this day, Salwa has continued in this effort, contributing what she calls a "creative solution for a long-standing problem...(that) alleviates the weight of escalating textbook costs and those brought on by other materials." (S. Ismail, personal communication, July 2, 2023)

Salwa has a long list of prior acknowledgements and accolades that she has earned, and accomplishments she has succeeded in, throughout the previous decade. In 2014, she was listed by E-campus News as one of the notable "11 leaders shaping the future of higher education." The following year, in 2015, she was named a 'Library Mover and Shaker' by the Library Journal for her work as a digital driver. She has chaired and been elected to several international and national committees on digital infrastructure, design, and services in libraries. In 2017, she was responsible for bringing to her institution the coveted National Digital Stewardship Residency (NDSR) program. Her passion lies in bringing technology and digital services to libraries and she considers herself an agent of research and scholarship for higher education, a platform for which she hopes to re-imagine a technological vision that she can then bring to the Library Sciences and

Information industries. Presently, in her role as Associate Librarian for Digital Initiatives, she not only continues in her work with the rapidly expanding Digital Lifecycle Program, but she also provides ongoing contributions to the library technologies division, with digital asset management, and other resource-sharing services in support of the University of California, Berkeley's mission of research, teaching, and innovation.

The UC Berkeley Digital Collections are a unique and fairly massive assemblage of manuscripts, books, photographs and other images, newspapers, all manner of other historical material, and rare digitized special collections. Salwa shared a number of details with me regarding the Digital Collections. They *do* incorporate fragile media, including audio-video and newspapers. A substantial portion has been digitized but not the whole of the collection. Administrative metadata for the digitized items is included in the applicable record, though despite all fields being captured in the preservation metadata, not all fields are published online. Information about ownership, rights, reproduction methodology, audit trails, persistent identifiers, access and citation requirements are all embedded in the record metadata. Though the metadata fields vary based on the item and the collection. The content of the Digital Collections is maintained in several various ways: one, the access copy of the digital version is published in the online digital collection platform, then subsequently backed up but not preserved, through regular IT backups. This is a type of 3-2-1 strategy, however, the preservation copy of the digital version (that normally comes in either tiff, jpeg2000, or wav, form) is preserved in the digital preservation repository – called Merritt. For vendor-licensed electronic content, they utilize LOCKSS, or "Lots of Copies Keep Stuff Safe," a library-inspired, peer-to-peer network, created under the aegis of Stanford University, that reinforces an open-source platform to collect, preserve, and access online published material, with a primary goal digital preservation. (Rosenthal, 2000, p. 2) Salwa emphasized that the Digital Collections do operate with an explicit digital preservation strategy that helps ensure that all digital assets, along with their metadata, are normalized, (at least to the extent possible,) and subsequently preserved for future access and use, should the need arise. Regular assessments and updates are conducted on the digital preservation policy. Though currently, any content that is digitized in-house, be it published online or not, born-digital items that are processed, and all vendor-licensed electronic content are included with the materials in the digital preservation workflows.

Copyright is taken very seriously and they've not encountered any problems thus far. Administrative metadata, including copyright, and other rights information is included with the digital assets as part of the ingest process. Email is included in the born-digital content type.

Additionally workflows along with tools, such as ePadd, are under exploration around this type of processing and preservation. The Collections' digital preservation strategy was defined in 2016-2017 initially based on the 2013 NDSA Levels of Preservation (LoP). Though, as Salwa detailed, as the tool has evolved, the workflows and processes have continually received ongoing upgrades to ensure updated content preservation levels are sustained. Currently, the preservation strategy maintains an assurance that all fully digitally materials processed for preservation consistently fall between a level 3 and level 4 for each of the functional areas. For content and control areas, the Collections hover around a level 3 based on capacity, technology, and staffing. An assessment was undertaken in 2020-2021, with an exploration of the analog formats that still require digitization within the archival collections, special collections, and other general collections being evaluated. Within the results of the 2020-2021 assessment, there were no overtly obvious shortfalls, but there *are* more materials than the Collections have the capacity to digitize immediately. Salwa relayed that the NDSA metrics are systematically used to record and visualize the assessment results of the Digital Collections though they have not thus far undergone an assessment using the DPC-RAM assessment metrics. She stressed that the team involved in the preservation of the Collections is aware that the NDSA assessment matrix is not the ideal for *all* collections and/or materials. It appears percentages are hard to define with regard to the amount of the Collections displayed online. Salwa revealed that despite a significant portion of the Collections being digitized and published online, there is also a large portion of digitized content that is not published online, but which is otherwise digitally preserved. The online materials are chosen via proposals from the subject librarian or curator, normally based on their research and use value. That said, there *are* materials chosen for online publications due to a fragile or damaged analog carrier. Within the online framework, the materials are organized under their respective library categories, and later cross-referenced in the metadata based on the various factors of subject area, geographic period, similarity to other collections, and multiple other attributes. Salwa contended that, in her opinion, the NDSA metrics are most relevant for institutions based in the US, as she felt the organization offers more integration and membership for domestic entities. She further denoted that her experience has led her to believe that the best assessment metrics tool is one that an organization's staff can understand and use without significant overhead, while also offering some measure of community - such as a listserv, conference groups...etc, wherein support and guidance is offered. She noted that the Digital Collections refrain from DPC-RAM, primarily due to the additional work and investments required of the staff. She *did* admit that she understands the DPC-RAM

assessment metrics, despite a notably more complex process, is capable of being equally, if not more so, beneficial as the NDSA metrics. She acknowledged that if staffing and expertise *does* exist to assist in the applicable areas, an exploration of the different functional areas of DPC-RAM that are not explicitly defined in the NDSA LoP matrix, would be extremely useful for an organization. Nonetheless, she reiterated that, in her professional opinion, the best digital preservation matrix tool is one “that an organization can easily and readily apply to its content to ensure their materials are being digitally preserved adequately and qualitatively.” (Ismail, personal communication, July 2, 2023) I tend to vehemently agree.

Below I’ve outlined what I believe are the most closely applicable factors for the individual categories of both the NDSA and DPC-RAM assessment metrics, inasmuch as I was able to infer. I found the following to be greatly challenging, first and foremost because after a short correspondence with Salwa, (that admittedly *did* relay a great deal of information,) she nevertheless informed me that she would subsequently be unavailable for any further communications due to her upcoming two-month vacation. Investigation of the Digital Collections’ website, and several notable documents were particularly useful. Most of all, I garnered a great deal of information from the 2020 ‘Phase Two’ report of the [University of California Digital Preservation Strategy Working Group](#). Additionally Salwa *did* mention that the work outlined at the end of the [Digital preservation leadership group: Phase 2.5 report](#) had been completed as scheduled. She also mentioned that many of the factors involved in the previous NDSA metric assessment were documented in the 2022 publication [Merritt digital preservation repository policies and user guidelines version 2.0](#). I found that not *all* categories involved in the DPC-RAM assessment were contained within the aforementioned document. However, I located and researched a multitude of other sources and in doing so, discovered a significant amount of information related to the dozens of individual respective metrics for the Digital Collections – an entity with a repository that is apparently governed with the overarching goal of allowing UC Berkeley to “to manage, archive, and share its valuable digital content.” (Lopatin, 2020)

Comparing the metrics is a formidable task. The below was summarized from a review of multiple sources, all of which are individually detailed below the subsequent paragraphs, in a following NDSA and DPC-RAM model detail. But Salwa *did* mention the prior 2020-2021 NDSA assessment resulted in a level between 3 and 4 for the categories of storage, integrity, and metadata, with a score closer to Level 2 for control and content. I am not exactly certain of why this is, but I imagine it may be due to the fact that the strategy makes a substantial effort where storage and

metadata area concerned. In observing the various factors, I noted more than one copy is stored at a geographic location with a different disaster threat than that at UC Berkeley. Storage diversification is also maximized, ostensibly to avoid single points of failure, and as can be inferred from the various factors I detailed following these last two paragraphs, there does appear to be a plan in place that addresses storage hardware, software, and media obsolescence. Integrity, in my opinion, fell closer to a Level 3 as there were replication sites present to implement actions to ensure integrity and authenticity. There is also significant cloud storage and ongoing verification efforts with a significant reliability. They maintain an audit process that's independent of local internal methods which provides external verification of all content. Though one of their cloud service providers, Glacier storage, was documented as not externally validated due to financial constraints. Metadata seems to hover between a Level 3 and Level 4 as the dash submissions are accompanied by DataCite 4.0 metadata, offering an interface enforced with only two of DataCite's mandatory elements: creator(s) and title. There appears to be one optional element: an abstract. The other elements involved, such as identifier, type, publisher, and publication date, are all programmatically added. Other elements, including funder, keywords, methodology, usage notes, related datasets/publications, and location may be supplied but remain optional. But as Salwa mentioned, information about ownership, rights, reproduction methodology, audit trails, persistent identifiers, access and citation requirements are all embedded in the record metadata. There does appear to be a policy in place mandating reasonable efforts be provided to manage content with the highest level of preservation assurance in mind, further noting this includes authoritative and comprehensive metadata. The element of Control was most challenging, as there was *very* little information that directly addressed this issue. Yet I took into account that the intellectual control of all UC Library material is given preservation responsibility, and that curatorially designated control must be established for either a restricted set of individuals for authenticated access and use or for unconstrained use by the public. Combined with Salwa's earlier comments, I felt that a designation of between a Level 2 and Level 3 was appropriate. Content was a further topic that little to no literature directly addressed. But I did discover that relationships between the Digital Library and content creators was encouraged with risk obviously mitigated by the storage of multiple copies across various storage platforms. They employ an extensive range of storage types located within multiple different geographic locations to reduce the risk of data loss. Up to three copies of data is stored within higher-education and other nonprofit centers in Texas, Maryland, and San Diego. This, of course, reduces the threat caused by

natural disasters as well as – to an extent - that from obsolescence. For the DPC-RAM model, Organization seemed to be a factor with which the Collections is struggling. As of 2022, they noted problems with various barriers to success and a lack of necessary actions required to enforce preservation structure. For Policy and Strategy, they seemed to be doing considerably better. They do appear to have an incredibly detailed and reinforced digital preservation policy. Despite being unable to acquire it specifically, I was able to surmise that a network of procedures and policies surrounding security, redundancy, independence, back-up and administrative procedures is currently in place. Though it seems more support is needed to flush it out completely. Legally speaking, I was able to locate almost nothing of relevant information. Though I noted that the specific framework in use does mandate that all laws, guidelines, regulations, policies surrounding ethics, and best-practices be consistently observed. For IT Capability, there was less information than I initially imagined but what I *was* able to find seem to suggest that expertise is available in nearly every applicable area, though it doesn't appear to always be used as needed. The most recent assessment seems to demonstrate that further attention and organization is needed for an ideal IT structure. For the area of 'Continuous Improvement,' I was not able to locate any information that I felt was genuinely applicable. But as I later note, given Berkeley's reputation, and the people at its disposal, and its ongoing official goal of development, I believe there's a high probability that they maintain continuing efforts for future improvements. Where Community is concerned, there is of course a UC-wide Digital Preservation Leadership Group composed of many professionals from multiple various backgrounds. Additionally the actual policies were originally created in a partnership between Berkeley, the University of Michigan, University of Illinois, the Society for American Archivists, and the American Library Association. It's to be assumed that the organization maintains continuing associations with these, and likely many other, related societies, associations, non-profits, other universities, and professional entities. For the Service Capability categories, I felt that, of the small amount of information I was able to uncover on Acquisition, Transfer, and Ingest a Level between Basic and Managed was warranted. It's quite possible, and likely, that other actual details I was unable to acquire provide more information that could relegate the category to a higher resultant level. But seeing as how Amazon FSx ingest services are currently listed as the body providing intermediate storage space, in addition to the fact that the team is still in the evaluation process for the overall benefit of the current service, I felt the aforementioned levels were appropriate. I did discover that fixity checks are regularly completed to catch ingest and replication errors, as well as other similar problems. Bitstream Preservation

appeared to be an area where they are doing well. Both of the applicable repositories, Chronopolis and Merritt, provide bit-level preservation through a regular procedural system. On top of that, the bit-level preservation the Collections *do* maintain ensures that each file in the repository remain static over time. But the entire repository undergoes regular fixity-checks via continual cryptographic verification put in place to uncover and rectify any bit-level corruption. Similar to the NDSA assessment, Content Preservation was a section I had ongoing problems with, but I did feel that both within that category, as well as, within the Management of Metadata, the same factors configured into the final resultant level. Finally, for the Discovery and Access category, I felt that the multiple policies, procedures, and many points of attention given to the topic of access mandated a final score, of the very least, between a Level 3 and Level 4.

Given Salwa's mention that the work laid out in the Phase 2.5 report is now complete, it would appear that there have been *significant* improvements to all of the previous categories. Along with the individual factors and the specific sources associated with all my aforementioned results, I've also provided a summary of that work and the results it provided. Following that there is an Appendix detailing components for UC IS-3 requirements that seem to denote very specific legal, ethical, security, and other related risks. Overall I would say that the UC Berkeley Digital Collections is an immense, quite overwhelming, but nonetheless, incredibly valuable library of singularly unique, interesting, and one-of-a-kind materials. Maintaining its preservation is incredibly important and as of the present time, it *does* appear that a very qualified team of professionals is in place to do just that.

NDSA Model:

Storage:

Level 4 – Sustain Content

Have at least one copy in a geographic location each with a different disaster threat
Maximize storage diversification to avoid single points of failure
Have a plan and execute actions to address obsolescence of storage hardware, software, & media

"Once the relevant metadata has been created and packaged with the preservation object, copies of the preservation objects are generated and placed in storage environments that are geographically dispersed and technologically distinct in order to reduce the risk of loss." (Schaefer, Chodacki, & Ismail, 2020, p.7)

Merritt relies on a primary strategy of replication to ensure the long-term integrity of managed data. All data is replicated to at least two geographically distributed locations and two heterogeneous technology stacks,

currently at the San Diego Supercomputer Center (SDSC) which uses the OpenStack Swift platform, and Amazon AWS, which uses S3 and Glacier. (CDL, 2018, p. 19)

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 Amazon AWS configuration information. (CDL, 2018, p. 31)

There are two collection management systems, Nuxeo and ArchivesSpace...Nuxeo is a digital asset management system (or DAMS) to create and manage object-level metadata and content files (such as images, texts, audio, and video). Nuxeo also supports workflows for publishing digital objects and for depositing digital objects into Merritt, for preservation. (OAC & Calisphere, 2021)

The Chronopolis network provides three copies of data stored at nonprofit, higher-education based data centers across the United States (San Diego, Austin, TX, and College Park, MD). (Schaefer, Chodacki & Ismail, 2020, p. 26)

"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)...and Wasabi Hot Cloud Storage is used as preservation storage for an additional object copy and is located on the East coast (Virginia)." (UC3, 2022, p.7)

Integrity

Level 4 – Sustain Your Content

Verify integrity information in response to specific events or activities

Replace or Repair corrupted content as necessary

Individual replication sites also implement internal mechanisms for ensuring integrity and authenticity. Merritt relies on SDSC's UC private cloud service and Amazon's AWS S3 and Glacier commercial cloud services for preservation storage, which are federated through Merritt's storage broker architecture. SDSC Cloud Storage implements an OpenStack Swift object store with three internal replicas on independent storage arrays, with ongoing MD5 verification of the validity of the replicas. AWS S3 and Glacier cloud storage maintain internal replicas and are designed to provide eleven 9's of reliability (99.9999999%) and sustain the simultaneous failure of any two internal copies. Merritt's Audit process provides external verification of all content independent of local internal methods at SDSC and S3. 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. (CDL, 2018, p. 16)

"Internally, the SDSC Swift cloud makes use of three independent replicas and its own internal digest-based auditing and self-healing capabilities. The S3 service description strongly implies that it also relies upon three independent replicas spread across availability zones with attendant internal fixity auditing and self-healing, claiming a 99.999999999% degree of durability." (CDL, 2018, p.19)

Control

Level 2

Document the human and software agents authorized to read, write, move, and delete content and apply these

"All digital assets under the intellectual control of a UC Library unit, for which the library assumes a preservation responsibility." (Peltzman & Schaefer, 2021, p. 7)

"...curatorially-designated access controls...permit designation for 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." (UC3, 2022, p. 3)

Metadata

Level 3

Determine what metadata standards to apply
Find and fill gaps in your metadata to meet those standards

Dash submissions are accompanied by DataCite 4.0 metadata. The Dash submission interface enforces the specification of only two of DataCite's mandatory elements: creator(s) and title, and one optional

element: abstract; the other mandatory DataCite elements ~ identifier, type, publisher, and publication date ~ are added programmatically. Other optional elements ~ funder, keywords, methodology, usage notes, related datasets/publications, and location (point, bounding box, or place name) ~ may also be supplied, and their use is strongly encouraged. (CDL, 2018, p. 18)

"This policy (mandates) reasonable efforts to provide managed content with the highest level of preservation assurance that is...accompanied by authoritative and comprehensive metadata, the availability of appropriate tools, and other organizational priorities." (UC3, 2022, p.4)

Information about ownership, rights, reproduction methodology, audit trails, persistent identifiers, access and citation requirements are all embedded in the record metadata. (Ismail, personal communication, July 2023)

Level 4
Record preservation actions associated with content and when those actions occur

Content

Level 2
Verify file formats and other essential content characteristics

Build relationships with content creators to encourage sustainable file choices
It is important to recognize that risk is mitigated by intentionally storing multiple copies of preserved content across a variety of storage platform types. Using a broad range of storage types at multiple geographic locations protects data and significantly reduces the risk of data loss. This robust approach to preservation storage provides genuine assurances that data can survive unexpected misfortunes (natural disasters, political instabilities, information warfare, operational malice, etc), but comes at a

monetary cost. (Schaefer, Chodacki & Ismail, 2020, p. 25)

The Chronopolis network provides three copies of data stored at nonprofit, higher-education based data centers across the United States (San Diego, Austin, TX, and College Park, MD). (Schaefer, Chodacki & Ismail, 2020, p. 26)

DPC-RAM Assessment

Metrics: The Model

A - Organizational viability

Governance, organizational structure, staffing and resourcing of digital preservation activities.

2 - Basic Digital preservation activities are supported & resourced at a basic level within the organization, for example:

- There is some engagement from senior management.
- Staff have assigned responsibilities and the time to undertake them.
- A budget for digital preservation has been allocated (may be time-limited).
- Staff development requirements have been identified.

"Based on the DPS WG's review of established practices at the individual campuses and department levels, insufficient staffing, ineffective organization, and a lack of training were the most significant barriers to progress ." (Schaefer, Chodacki, Ismail, et. al. 2020, p. 3)

"During the interviews, our foremost observation was the critical need to organize digital preservation activities across different departments...preservation efforts are frequently not aligned with the existing organizational structure of the library, and there is friction and confusion as preservation roles are applied in different

groups." (Schaefer, Chodacki, Ismail, et. al. 2020, p. 18)

B - Policy and strategy

Policies, strategies, and procedures which govern the operation and management of the digital archive.

3 – Managed - The organization has a comprehensive and managed suite of policies, strategies and procedures, for example:

- The digital preservation policy/strategy is aligned with other organizational policies and is reviewed according to an agreed schedule.
- Policy and procedure takes into account any relevant ethical issues.
- A suite of documented processes and procedures for managing, and providing access to, content within the digital archive exists.
- All relevant staff are aware of digital preservation policies, strategies and procedures.
- Knowledge of current and future use cases for content informs policy and procedure (for example on collecting, preservation approaches, metadata and access).

"Policies provide essential guidance and can help establish the appropriate organization and governance of digital preservation activities, It was agreed across the board that policies were desirable to guide practice and allow for administration at the appropriate level. In 2019, it was determined that an outline of core function-specific digital preservation practices, is in place, but further attention is needed. Additionally, without staff in place to implement policy as action, policies by themselves are ineffective at instilling change. Because digital preservation is one piece of many processes across the library, there are many that tangentially relate to, and could be strengthened by, more direct support for digital preservation work. In order to move forward in establishing new policies and

reviewing existing ones, a formal assessment of the current policy framework at the campus and system level should be undertaken." (Schaefer, Chodacki & Ismail, 2020, p. 22)

Both of the digital preservation systems, Chronopolis and Merritt, provide bit-level preservation, achieved through a combination of policies and procedures that govern data security, storage, redundancy, independence, and backup for disaster recovery." (Schaefer, Chodacki, & Ismail, 2020, p.6)

C - Legal basis

Management of legal rights and responsibilities, compliance with relevant regulation and adherence to ethical codes related to acquiring, preserving and providing access to digital content.

2 – Basic Basic management of legal rights and responsibilities relating to digital content is carried out, for example:

- Key legal rights and responsibilities, together with their owners, have been identified and documented.
- Templates exist for necessary legal agreements and licences.
- Relevant codes of conduct relating to professional ethics are adhered to.

"...(before publication,...)All applicable laws, regulations, policies, ethical concerns, and disciplinary best practices regarding the creation and acquisition of...content, including obligations regarding intellectual property rights, privacy, IRB review, and accepted norms of scholarly discourse, have been observed and followed and the (Digital Library) has been assigned 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 (is) subject to curatorially-designated access controls...permit designation for 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."(UC3, 2022, p. 3)

**I was unable to find any information at all on the legal representation the Digital Collections has at their disposal, although Salwa commented on how seriously copyright is taken and from that, I can surmise that all issues with any type of legality are given the utmost attention.*

D - IT capability

Information Technology capabilities for supporting digital preservation activities.

4 – Optimized - The organization has access to proactively managed IT facilities that are continually evolving and improving, for example:

- An enhanced level of IT support is available to the digital archive
- IT demonstrates good understanding of, and engagement with, digital preservation issues.
- Digital preservation requirements are taken into account when sourcing new IT systems.
- A detailed roadmap exists for future development of IT systems.
- Potential new tools and systems are proactively identified and tested.

"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." (UC3, 2022, p. 3)

"A lack of expertise (was found)... among current staff...(and) also a mutual lack of awareness among...other campuses' digital preservation personnel expertise, initiatives, and activities, resulting from and reinforcing a lack of communication and collaboration....Likewise, the existing distributed expertise in digital preservation

could be leveraged more efficiently by creating a shared service model for digital preservation serving all UC libraries." (Smith, Chodacki, Elings, et. al., 2019, p. 4)

"A lot of expertise (exists) but there is overlap and...we are not collaborating on digital preservation. The present distributed expertise in digital preservation would be used much more efficiently by creating a digital preservation shared service model which could serve (the entire) (digital library). (Smith, Chodacki, Elings, et. al., 2019, p. 13)

E - Continuous improvement

Processes for the assessment of current digital preservation capabilities, the definition of goals and the monitoring of progress.

4 – Optimized - The organization undertakes continuous process improvement, with proactive management, for example:

- Certification/external review has been achieved and is maintained as appropriate.
- Recommendations for improvement have been acted upon.
- Goals and roadmap are reviewed periodically.

**I am uncertain about these points, as despite encountering peripherally related topics throughout the documentation and literature, the only concrete mention of 'improvement' I discovered was:*

"The Merritt Team undertook a number of initiatives to improve the sustainability of the system itself to ensure it can be managed and maintained by the current team as well as by future team members." (CDL, n.d, Merritt System, p.2)

There was no information that directly related to the policies, efforts, and/or official goals for ongoing future improvements. But I can infer from both Berkeley's reputation, the people it has at its disposal, and its normal lean toward evolving developments that there is a good chance plans addressing continuous improvement are in place.

F - Community –

Engagement with and contribution to the wider digital preservation community.

3 – Managed Engagement with the wider digital preservation community is supported and managed, for example:

- Relevant networks and communities have been joined.
- An active role is taken in the digital preservation community.
- Expert advice on digital preservation can be accessed as appropriate.
- Successes and lessons learned from own work is shared with the community.

"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." (UC3, 2022, p.4)

Our principles policies were **developed in part** with reference to and reliance on those of other academic libraries, memory institutions, and professional organizations, including **University of Michigan, University of Illinois, the Society for American Archivists, American Library Association, and others.**) (UC Berkeley Library, 2023, 'About,')

UC Digital Preservation Strategy Working Group Author Info: Sibyl Schaefer (UC San Diego,) John Chodacki (California Digital Library,) Salwa Ismail (UC Berkeley,) Greg Janée (UC Santa Barbara,) Eric Lopatin (California Digital Library,) Charlie Macquarie (UC San Francisco,) Kevin Miller (UC Davis,) Erik Mitchell (UC San Diego,)

Shira Peltzman (UCLA,) Adrian Petrisor (UC Irvine,) Chrissy Rissmeyer (UC Santa Barbara,) Edson Smith (UCLA,) Roger Smith, (UC San Diego,) Sarah Troy, (UC Santa Cruz,) (Schaefer, Chodacki & Ismail, 2020, Author Page)

**I am uncertain about this specific section, as there was very little documentation and/or literature directly relating to the effort of the Berkeley Digital Library to remain in working partnerships with other related organizations. However I can imagine that, as reputable and giant as they are, there is a natural gravitation toward doing this on a regular basis.*

Service capabilities

G - Acquisition, transfer and ingest

Processes to acquire or transfer content and ingest it into a digital archive.

2 – Basic - The organization has implemented a basic process for acquisition, transfer and ingest, for example:

- **A documented ingest process exists.**
- Basic guidance for donors, depositors and record creators is available where appropriate.
- Documentation and metadata is sometimes received or captured as part of the acquisition or transfer process.
- A documented process exists for selecting and capturing digital content where appropriate (for example for web archives, email archives, digitized content, records within an EDRMS)
- Some content is appraised as part of a manual process in line with relevant policies.
- **A working area (physical or virtual) is available** for pre- ingest and ingest activities (for example to carry out virus checking and file identification).

The Ingest and Storage services of Amazon FSx are being used for OpenZFS as an

intermediate storage space during the ingestion process rather than AWS EFS. Though the team is still gauging overall throughput benefits of ZFS, we're encouraged by the initial results we've observed in terms of network bandwidth and speed. (Lopatine, 2022)

"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." (UC3, 2022, p. 3)

H - Bitstream preservation

Processes to ensure the storage and integrity of digital content to be preserved.
should be authorized to access the content.

3 – Managed - The organization stores content in a managed way consistent with preservation good practice for replication and integrity checking. For example:

- Content is managed with a combination of integrity checking and content replication to one or more locations.
- Decisions on the frequency of integrity checking and the number of copies held take into consideration risks, value of the content and costs (both financial and environmental).
- Content failing integrity checks is repaired.
- Authorizations to access the content by staff are enforced and documented.
- Tests are routinely carried out to verify the effectiveness of backups, replication and integrity checking.

Both of the digital preservation systems, Chronopolis and Merritt, provide bit-level preservation, achieved through a combination of policies and procedures that govern data security, storage, redundancy, independence, and backup for disaster recovery. " (Schaefer, Chodacki, & Ismail, 2020, p.6)

"The primary preservation strategy (maintained here)...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....Through object versioning...a complete change history of managed content (is maintained) as it may evolve over time. All undergo routine fixity-checks 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." (UC3, 2022, p. 2-3)

I - Content preservation

Processes to preserve the meaning or functionality of the digital content and ensure its continued accessibility and usability over time.

4 – Optimized The organization takes a proactive approach to prioritize and mitigate preservation risks to ensure content is accessible over time, for example:

- Risks to specific file formats or types of content held are well understood.
- A rigorous preservation planning process identifies appropriate preservation actions for risk mitigation.
- Decisions on whether to enact preservation actions take into account risks, value of content, costs (both financial and environmental) and use cases.
- Format migrations, normalizations, emulation and other preservation actions are implemented in accordance with preservation plans.
- Quality control is in place to assess (and record) the outcome of preservation actions, ensuring that the meaning and/or functionality of the content has been retained as required.

- Digital content and metadata are version controlled where appropriate.

It is important to recognize that risk is mitigated by intentionally storing multiple copies of preserved content across a variety of storage platform types. Using a broad range of storage types at multiple geographic locations protects data and significantly reduces the risk of data loss. This robust approach to preservation storage provides genuine assurances that data can survive unexpected misfortunes (natural disasters, political instabilities, information warfare, operational malice, etc), but comes at a monetary cost. (Schaefer, Chodacki & Ismail, 2020, p. 25)

"(Preservation) relies on internal and external service providers for primary and replication storage in its preservation system as well as its compute hosts." (UC3, 2022, p.6)

"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)....and Wasabi Hot Cloud Storage is used as preservation storage for an additional object copy and is located on the East coast (Virginia)." (UC3, 2022, p.7)

J - Metadata management

Processes to create and maintain sufficient metadata to support preservation, discovery and use of preserved digital content.

3 – Managed - The organization has implemented a managed process to create and maintain metadata for preservation, discovery and use, for example:

- **Appropriate metadata standards are identified.**
- Internal guidance and controlled vocabularies are in place to ensure consistency of metadata entry.

- Persistent unique identifiers are assigned and maintained for digital content.

● **Structural relationships between the data and metadata elements that form a particular digital object are maintained.**

"Once the relevant metadata has been created and packaged with the preservation object, copies of the preservation objects are generated and placed in storage environments that are geographically dispersed and technologically distinct in order to reduce the risk of loss." (Schaefer, Chodacki, & Ismail, 2020, p.7)

Merritt relies on a primary strategy of replication to ensure the long-term integrity of managed data. All data is replicated to at least two geographically distributed locations and two heterogeneous technology stacks, currently at the San Diego Supercomputer Center (SDSC) which uses the OpenStack Swift platform, and Amazon AWS, which uses S3 and Glacier. (CDL, 2018, p. 19)

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 Amazon AWS configuration information. (CDL, 2018, p. 31)

There are two collection management systems, Nuxeo and ArchivesSpace...Nuxeo is a digital asset management system (or DAMS) to create and manage object-level metadata and content files (such as images, texts, audio, and video). Nuxeo also supports workflows for publishing digital objects and for depositing digital objects into Merritt, for preservation. (OAC & Calisphere, 2021)

"This policy (mandates)...reasonable efforts provide managed content with the highest

level of preservation assurance consistent with form, structure, and packaging of the content, the degree that it is accompanied by authoritative and comprehensive metadata, the availability of tools, and other organizational priorities." (UC3, 2022, p.4)

Information about ownership, rights, reproduction methodology, audit trails, persistent identifiers, access and citation requirements are all embedded in the record metadata. (Ismail, personal communication, July 2023)

K - Discovery and access

Processes to enable discovery of digital content and provide access for users.

3 – Managed - The organization has implemented a comprehensive, managed discovery and access process (where access rights permit), for example:

- **Basic resource discovery exists for all digital content.**
- **Full text search is available for some digital content.**
- **Rights information is displayed** and access managed by the system where possible.
- **Reports can be generated about user access to digital content.**
- Access systems are updated to reflect feedback from the user community.
- Resource discovery information is available to users in accessible formats.
- Established access use case for mass extraction of all digital content during invocation of an exit strategy.

"UC's accessibility policy...promotes an accessible IT environment...to help ensure that as broad a population as possible may access, benefit from, and contribute to the University's electronic programs and services... and the (Digital Library) has been assigned 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 (is) subject to curatorially-designated access controls...permit designation for 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." (UC3, 2022, p.3)

"All access log information and other personally-identifying evidence of use is collected and dispositioned in a manner consistent with the CDL privacy policy" (UC3, 2022, p. 4)

"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 value and provide a higher level of assurance of its ongoing availability and usability." (UC3, 2022, p. 6)

Work Cited as Completed as of 2022

1. ADMINISTRATIVE.

In March 2021 the Digital Preservation Leadership Group (DPLG) selected a Chair and Vice Chair; the Chair and Vice Chair then appointed a Project Manager, who scheduled regular recurring bi-weekly meetings and established a system for communication. DPLG appointed and charged a Training Program, Assessment Matrix, and Collaborations subgroups responsible for leading that work. Additional activities included:

- A review of the charge and agreement on the shared interpretation of the DPLG's purpose, goals, and priorities
- A community agreement was established that outlined expectations around group participation and engagement

- A listserv and wiki for DPLG communication was set up
- Protocols and expectations for regular website updates were determined

● **In July 2021** the DPLG submitted their first annual report to Direction and Oversight Committee (DOC,) the Council of University Librarians (CoUL,) and the UC community about the activities of the DPLG, and published key information about the DPLG, including membership, charge, and priorities. Additional activities also included:

- Outreach to publicize key information about the DPLG, including a description of activities currently planned and underway
- A webinar about the outcomes of Phases 2 and 2.5 to the UC community.

● **In January 2022** the DPLG requested financial and resource allocation from CoUL for Training Program and/or Assessment Matrix deliverables.

- In July 2022 the DPLG submitted their second annual report to DOC, CoUL and the UC community about the activities of the DPLG, publishing key updates on the website, and a review on its value within the UC Libraries Advisory Structure. Additional activities also included:
- Reassessment, re-prioritizing, and re-outlining of work plans for future and ongoing work

2. TRAINING

In summer 2021, training was developed, enhanced, and a systemwide digital preservation training program was established. The intended audience was the existing preservation staff as well as administrators, team leads, and unit/department heads with digital preservation either fully or partly within their portfolios.

● **In July 2021**, training began on the overall goals and learning outcomes.

A subgroup created and submitted a proposal using DOC's project proposal template.

Additional activities included:

- A project kick-off meeting that developed a work plan
- A literature review & scan of existing educational modules, resources, and training materials
- A UC-wide survey of existing preservation staff that also included administrators, team leads, or unit/department heads with digital preservation either fully or partly within their portfolios
- A report was prepared outlining the DOC project and the contributions to DPLG's annual report

● **In January 2022** the Training Program subgroup identified target dates for training; applied for funding to secure consultants and hire graphic designers and training materials were developed. Additional activities also included:

- Applications to external grants for educational materials
- Appropriate consultants were identified
- Announcements and Save the Dates were distributed to attendees
- Consultations with campus staff were planned to address specific training needs

● **In July 2022** the Training Program subgroup delivered the initial training modules and began the assessment process. Additional activities included:

- A follow-up assessment survey was conducted
- Subsequent training programs were developed, with appropriate adjustments to form and content
- A report was prepared on the DOC project

3. ASSESSMENT

At the same time, a draft of a refined assessment matrix was outlined to assist campuses with determining the appropriate level of stewardship for a given set of digital

materials. This involved representation from across UC and was done with respect to each individual library's collecting policies and practices.

● **In July 2021** an Assessment Matrix subgroup also began laying the groundwork for a proposal using DOC's project proposal template. This included:

- A project kick-off meeting for developing the work plan
- A review of Phase 2 stakeholder interviews to ascertain existing collecting policies and practices and identify their significant characteristics
- A list was constructed using a set of predefined categories of digital materials currently being collected and those for potential collection
- A literature review of pertinent assessment tools was conducted to assist with digital preservation decision-making and resource allotment
- A report was prepared outlining the DOC project

● **In January 2022** the Assessment Matrix subgroup drafted a matrix and circulated it for review. Additional activities also included:

- A research study was conducted to assess levels of stewardship and define a set of preservation activities, that were then mapped into tiers based on the resources required to perform them
- Accompanying documentation was drafted to support the application of the matrix
- A review of the matrix was conducted with the full DPLG membership.
- Campus staff tested the effectiveness of the draft matrix on a limited number of collections

● **In July 2022** the Assessment Matrix subgroup produced a final draft of the matrix. After the completion of the final draft, the subgroup then:

- Distributed the matrix to ensure that all relevant materials and associated documentation were publicly accessible
- Training was provided on matrix use for all existing digital preservation staff but also with many administrators, team leads, and unit/department heads with digital preservation either fully or partly within their portfolios
- Materials and associated documentation were made publicly accessible
- A report was prepared on the project outcomes for DOC

4. COLLABORATIONS.

Frameworks have now been established for administering and facilitating cross-campus engagement with external collaborators and consortial partnerships, especially with regards to grant opportunities.

● **In March 2021** the DPLG established a small subgroup to document existing practices for cross-campus grants and UC-wide partnerships and memberships. The subgroup created and submitted a proposal using DOC's project management template.

● **In July 2021** the subgroup put forth recommendations for the DPLG to adopt regarding how to identify grant and partnership opportunities, collaborate across the campuses, determine the appropriate administration of resources, and manage ongoing reporting requirements.

Recommendations were also included for mechanisms to review the processes at appropriate intervals.

● **In September 2021** the framework for administering and facilitating cross-campus engagement was approved by the full DPLG and DOC.

Future and Ongoing Work:

In addition to special projects and/or grant opportunities, the following activities are recommended and will need to be

accomplished in the future: ●**ECONOMIC MODELS.**

Analysis of available economic models that quantify and assess both costs and benefits, and establish which can be applied to the UC system. (DP WG Phase 2 Report Recommendation 2C)

●**CAMPUS-BASED STRATEGIES.** The DPLG should support and work with designated individuals/groups at each campus to analyze their current policies related to digital preservation, compare them against established frameworks, and determine where gaps exist. (DP WG Phase 2 Report Recommendation 3A&B)

●**PROTECTED DATA.** The best path forward should be investigated for addressing HIPAA compliance, as well as the preservation of other kinds of sensitive digital information, including FERPA-protected and other confidential material. (DP WG Phase 2 Report Recommendation 2D)

●**REPORTING.** Report to DOC, CoUL and the UC community annually about the group's activities. Update website and provide periodic status briefs, as requested. (Schaefer, Chodacki, Comerford, et. al, 2021, p. 9-11)

Appendix D	
Data Classification Matrix elements	
Charge sub-bullet: Identify necessary components for a matrix or rubric to classify digital information/content held by the UC Libraries, in order to assess how it maps onto UC Information Security requirements. This matrix would ultimately be completed by the future UCLAS Digital Preservation group, and may, if appropriate, be approved by campus CISOs or other appropriate campus representatives.	
Components of a Matrix map for UC IS-3 requirements	
Confidential or Sensitive	Does the material have any Attorney-Client or other privileged information, donor agreements, academic records with faculty student grading information, FERPA data? Does it contain staff or user information such as phone numbers, citizenship, income ta data, address, performance evals, confidential academic review records?
Notice-triggering information (HIPAA, GDPR, employee, patron, user data, etc.)	Do the material have any SSNs, Drivers Licenses, bank info, medical records, health insurance; Details – Does the material contain Social security number, Government issued identification numbers (Driver's license number, California identification card number, tax identification number, passport number, military identification number, or other unique identification number issued on a government document commonly used for identity verification), Financial account numbers, credit or debit card numbers, and financial account security codes, access codes, or passwords, Personal medical information, Personal health insurance information, Biometric data used for authentication purposes, including photographs used or stored for facial recognition purposes, A username or email address, in combination with a password or security question and answer that would permit access to an online account, Information or data collected through the use or operation of an automated license plate recognition system, Separate but related is personal information under the General Data Protection Regulation (GDPR)
Rare/unique	Is the material rare, unique, difficult to obtain or reproduce?
Copyrighted	Is the material protected by copyright?
Contractual restrictions	Does the material have any specific contractual obligations around access beyond IP or campus restriction?
Ethical concerns	Does the material have any ethical concerns? Around NAGRPA and the CARE principles. Could the information in this material lead to: harm or exploitation of people, resources, or knowledge?
Privacy concerns	Does other content fall under one or more of these privacy torts? (e.g. nudity, illegal activity, private personal information -birthdate, death date, marriage date etc.)
Access restriction (controlled but not confidential/unclassified)	Does the material need to be accessed through some form of authentication (IP, campus-user access, identify management, or authentication of any kind etc.)?
PR risk (reputational damage)	Would the loss or leak of this material or content in this material create a PR debacle for the organization? Will such a loss or leak lower the reputation of the organization?
Access/Retrieval	Does this material need to be accessed often by users? Will it need to be retrived frequently or occasionally?
Business controlled data (institutional information)	Does the material contain data and information created, received and/or collected by UC relating to the activities or operations of the university or organization?
Shared-Fate Information	Would material/data or system compromise cause further and extensive compromise to multiple (even unrelated) sensitive systems? e.g.: e.g., enterprise credential stores such as the username credential database; Domain Name Service (DNS).

(Schaefer, Chodacki, Comerford, et. al, 2021, p. 13)

Bibliography:

Asimov, N. (June, 2012). "UC Berkeley library's next chapter may be cuts." *University Council AFT*.
<https://ucaft.org/content/uc-berkeleys-libraries-next-chapter-may-be-cuts>

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>

Chan, L., Miller, B., & Hamed, M. (18, Jan., 2023) *Library impact research report: Open access publishing: A study of UC Berkeley faculty views and practices*. Washington, DC: Association of Research Libraries. <https://doi.org/10.29242/report.ucberkeley2023>.

Digital Scriptorium. (n.d) *About DS*. <https://digital-scriptorium.org/about/about-ds/>

Early American Sources. (n.d.). *United States archives*.
<https://www.earlyamericansources.org/united-states-archives>

Grappone, T. Elings, M., & Smith, E. (2019). *Building a digital preservation strategy across a broad university system*. https://www.cni.org/wp-content/uploads/2020/01/CNI_buildingdig_smith.pdf

Greenstein, D., Lawrence, G. et. al. (1, Jan., 2003). "New models of library service: Deep resource sharing and collaboration at the University of California." *Digital Libraries: A Vision for the 21st Century*. <https://doi.org/10.3998/spobooks.bbv9812.0001.001>

Greenstein, D. and Thorin, S. (2002). *The digital library: A biography*. Digital Library Federation Council on Library and Information Resources. <https://www.clir.org/wp-content/uploads/sites/6/pub109.pdf>

Haugan, T. (19, Aug., 2021). *The UC Berkeley library is leveling the playing field, one textbook at a time*.
<https://www.lib.berkeley.edu/about/news/e-reserves>

Heidelberg. (23, Mar., 2021). "Springer Nature and UC Berkeley library sign new open access book partnership." *Springer Nature*. <https://librarytechnology.org/pr/26003>

Ismail, S. (16, Feb., 2022) *Archiving and preservation of unusual born-digital objects*.
<https://niso.cadmoremedia.com/Title/3fc56d7e-053c-434d-8d09-183eed8458e6>

Ismail, S. (25, May 2022). *One decision at a time: Reimagining decision-making for democratized technology projects*. <https://2022.code4lib.org/talks/One-Decision-at-a-Time-Reimagining-DecisionMaking-for-Democratized-Technology-Projects>

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/>

MacKie-Mason, J. (2021). *FY: 2020-2021 annual report*. UC Berkeley Library.

https://www.lib.berkeley.edu/sites/default/files/uc_berkeley_library_annual_report_2020-21.pdf

MacKie-Mason, J. & Samberg, R. (16, Apr., 2020). *UC Berkeley's Digital Lifecycle Program: Mass digitization of special collections for use and preservation*.

<https://www.ischool.berkeley.edu/events/2020/uc-berkeleys-digital-lifecycle-program-mass-digitization>

Markovich, A. (3, Mar., 2023). "UC Berkeley will close 3 libraries amid shrinking budget."

Berkeleyside. <https://www.berkeleyside.org/2023/03/03/cal-berkeley-library-closures-anthropology-physics-mathematics>

OAC, Calisphere, (24, Aug., 2021). *UC shared DAMS (Nuxeo) and ArchivesSpace*.

<https://cdlib.org/services/pad/ucdams/>

Peltzman, S. & Schaefer, S. (2021). *Advancing digital preservation across the UC*.

<https://d3smihljt9218e.cloudfront.net/lecture/38153/slideshow/835ff731ad95f9b2caf07910523dfdbf.pdf>

Rosenthal, D. & Reich, V. (18, June, 2000). *Permanent web publishing*. Proceedings of FREENIX Track: 2000 USENIX Annual Technical Conference.

http://www.usenix.org/events/usenix2000/freenix/full_papers/rosenthal/rosenthal.pdf

Schaefer, S., Smith, E., & Tashjian, H. (25, July, 2022). "Building a path forward to sustainable digital preservation: The genesis of digital preservation leadership across the UC system." CNI.

<https://www.cni.org/topics/digital-preservation/building-a-path-forward>

Schaefer, S., Chodacki, J., Comerford, K. et. al. (Feb., 2021). *Digital preservation leadership group: Phase 2.5 report*. https://libraries.universityofcalifornia.edu/wp-content/uploads/2021/04/DPS_Phase_2.5_Report_Feb2021.pdf

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>

Serra, A. & Elings, M. (10, January, 2022). "California land case files from the 1850s to be digitized." *Society of California Archivists Newsletter*, 181, 6-7.

<https://calarchivists.org/resources/Documents/Newsletter/2022Winter.pdf>

Shah, U. (2019). "LOCKSS, CLOCKSS, & PORTICO: A look into digital preservation policies."

Library Philosophy & Practice. <https://core.ac.uk/download/pdf/215162116.pdf>

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>

Tsung,S. &Ismail, S. (2014). *MOOC-ing up your institution's IR collections*.
[https://www.ala.org/lita/sites/ala.org.lita/files/content/LITA Concurrent Sessions 2014 rev 20141103c.pdf](https://www.ala.org/lita/sites/ala.org.lita/files/content/LITA_Concurrent_Sessions_2014_rev_20141103c.pdf)

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>

UC Berkeley Library. (2023). *Berkeley library facts and figures*. University of California, Berkeley.
<https://www.lib.berkeley.edu/sites/default/files/files/UCBLibraryFacts.pdf>

UC Berkeley Library. (n.d). *Information for visitors*. <https://www.lib.berkeley.edu/help/visitors>

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