University of California
CDL
California Digital Library

California Digital Library
The Second Decade

Prepared for the review of the CDL

By the California Digital Library

2010

Table of Contents

I. Introduction 3
Background: CDL Yesterday and Today 3
Looking Forward 4
II. Context and Value Proposition 5
The CDL's Role and Value Proposition 6
CDL Programs and Services 8
Collect 8
Access 11
Publish 11
Preserve 13
III. Prioritization and Decision-making 15
IV. Sustainability 18
Conclusion 24

California Digital Library: the Second Decade
September 2010

I. Introduction

Scholarly information is the lifeblood of the academy.[1] The faculty, students and staff of the University of California who embody the University's research, teaching, and public service mission both produce and consume scholarly information in the act of creating and sharing knowledge and rely on the prior record of human history, scholarship, and evidence to nourish and inform their work. This time-honored practice of academic knowledge production and exchange is in the midst of a transformative shift away from traditional forms and methods, toward what Charles Henry, President of the Council on Library and Information Resources, has described as a "collective relocation…to a digital environment for knowledge access, preservation, and reconstitution".[2] To put it simply, the future of scholarly information is digital.

This digital "relocation" – paralleling more general and equally pervasive cultural trends – is of profound significance for the University and its libraries. The digital environment is blurring the lines between knowledge creation and formal publication, radically changing the way scholars and students find, access, and use information, and creating new demands for effective curation of digital content of increasing volume and complexity to ensure that today's scholarship will not be lost to future scholars.

Digital libraries have emerged as a strategic institutional response to these evolving scholarly information needs, providing key services and resources for access to knowledge and for the stewardship of the University's intellectual assets. The California Digital Library (CDL) is internationally recognized as a leading example of this essential institutional role; a reputation acquired in little more than a decade of service to the University that is a tribute to the foresight of the University of California leaders through whose vision the CDL was conceived.

Background: CDL Yesterday and Today

Founded by the University of California in 1997, the CDL was charged as the principal University agent to develop services that could both respond to and anticipate the evolving scholarly information needs of the UC academic community. Specifically, CDL's mandate was to work in partnership with the UC libraries to:

1. enhance information resource sharing across the campuses;
2. create a scholarly information infrastructure;
3. collaborate across and among disparate information-intensive partners including other libraries, industry and museums;
4. transform scholarly communication (in part by becoming a "digital publisher");
5. use information technology to support the effective use and sharing of print materials; and,
6. create a framework for continuous planning and innovation.

The 1997 "Library Planning and Action Initiative" that named these goals and recommended the establishment of the CDL, framed them as part of an overall effort to create "a sustainable model or

models for the University Library System to accommodate changing funding, intellectual, service, collection development and technology environments."[3]

An ensuing UC libraries strategic planning initiative of 2004 and its 2005 progress report implicitly reaffirmed these goals and added to them several new ones. It called out the importance of enhancing resource sharing and influencing scholarly communication, including through the development of innovative publishing alternatives; and it identified "persistent access," a reference to the increasingly urgent need for digital preservation and curation, as an explicit strategic goal.[4] The CDL, in turn, has continued to reaffirm its role in meeting these goals and in providing "systems and services that can make the University's shared knowledge assets in any format readily accessible and available to every member of the UC community."[5]

Looking Forward

In its short twelve-year history, the CDL has established an unrivaled suite of information services[6] for the University and its partners and a reputation for operational and strategic excellence.[7] The CDL's services are used by every segment of the University and its constituencies. Our users represent many stakeholders and partners, including the UC campus libraries, museums and archives; faculty and researchers working directly with the CDL and through their campus libraries; cultural heritage institutions throughout the state; and California citizens and the general public.

But the CDL cannot rest on its laurels. As the digital transition unfolds, its contours are becoming clearer, as are its challenges and opportunities. These affordances come at a time of significant economic turmoil for the University as a whole. The University, as indeed all of academe, currently faces unprecedented financial challenges and an accompanying need to assess and, where appropriate, refocus its efforts.

In charting a way forward for the University, the UC Commission on the Future asks "How can the University of California best serve the state in the years ahead and maintain access, quality and affordability in a time of diminishing resources?"[8] This challenge demands of CDL a corollary question: "How can the *California Digital Library* best serve the University of California in the years ahead and maintain access, quality and affordability of scholarly information in a time of diminishing resources?"

It is gratifying that in attempting to answer that question, the Commission advises the University to "maximize the UC library system's capacity to support the University's research mission by: enhancing and developing data curation techniques; extending systemwide acquisition and sharing of resources; expanding accessibility of physical and virtual library space; and promoting systemwide scholarly publishing initiatives."[9]

Nonetheless, achieving these and other goals we may identify will depend on our ability to articulate a clear value proposition within the context of the digital information imperatives our institution faces, prioritize our efforts, and adapt CDL's funding requirements to the University's current capacity. The present document outlines the CDL's current efforts and approach to these questions.

II. Context and Value Proposition

The context in which the CDL operates includes its original and ongoing mission to help build UC's digital information infrastructure as well as the current financial constraints of the University. UC's information infrastructure, in turn, must serve the rapidly evolving practices of the global research and teaching enterprise as well as the particular manifestations of those practices within UC.

Two important early reports documenting the evolution of scholarship in light of digital and network technologies were the 2003 Report of the National Science Foundation (NSF) Blue-Ribbon Advisory Panel on Cyberinfrastructure and a 2006 complement from the American Council on Learned Societies (ACLS) on cyberinfrastructure in the humanities and social sciences.[10] Both reports describe emerging new forms of digital scholarship and the structures and tools needed to support them, including the integration of increased capacities for computation, content, and interaction,[11] and the use of "digital collections and analytical tools to generate new intellectual products."[12] A half dozen years later, digital scholarship is becoming more firmly embedded in the academic enterprise. Disciplinary examples are legion, ranging from the recent award by the National Endowment for the Humanities (NEH)'s Office of Digital Humanities of one million hours of supercomputing time for humanities research[13], to the Biomedical Informatics Research Network, pre-eminent among thousands of emerging scientific "collaboratories."[14]

These trends manifest themselves in specific needs and expectations of scholars and students at UC. Specifically, UC's scholars:

- **Are among the most intensive users of online information resources worldwide**
  Licensed online journals and databases at UC were used more than 47 million times in 2009. As the number of such resources has grown over the past decade, their usage has increased by an average of 11% per year; year after year, publishers and other online providers consistently report that the University of California ranks at or among the very top institutional users of their content across the globe.

- **Expect immediate, unmediated access to online research materials and data.**
  While UC library circulation transactions for print materials dropped by 50% from 2000 to 2008,[15] the use of the UC libraries' direct linking service, UC-eLinks, has grown at an average annual rate of 14% for the last six years. UC-eLinks redirected users from citations to online content and delivery services 900,000 times in one month alone in April 2010. Many of these transactions result from the increasing use of search engines such as Google, Google Scholar, and services such as PubMed in addition to traditional library catalogs and indexes.

- **Generate increased quantities and varieties of digital information.**
  In addition to publishing over 26,000 journal articles per year,[16] UC faculty members operate some of the world's most advanced telescopes, supercomputers, and other scientific instruments, producing the digital data from which today's and tomorrow's findings are derived. Traditional formats are being replaced or supplemented by new ones - for example, UC manuscripts that are peer-reviewed via blogs[17], and publications that depend upon online computer-generated spatial visualizations.[18]

- **Disseminate their results through alternative venues and in new formats.**
  Scholarly findings are no longer simply communicated via traditional publication channels and formats. Faculty have begun to share their work through the direct dissemination of original research data sets, 3-D visualizations, video content, blogs, and more. Private and public research funders increasingly demand that research results be made available for public access through disciplinary and institutional repositories.[19]

  The trend toward increasingly diverse and multifaceted models of scholarly research and communication is complicated by a challenging copyright landscape, prompting the need for new rights management mechanisms to catch up with the digital information environment.[20]

- **Are increasingly required to arrange for the management of research data.**
  In keeping with the trend toward 'data as publication,' in April 2010 the NSF announced that data management plans and follow through are now required as a condition of research funding.[21] The NIH has a long-standing similar requirement.[22] Compliance with these mandates has important consequences both for individual UC researchers and for UC's competitive advantage: the 10-campus UC system consistently earns two to three times more NSF research support than any other university system ($593 million in 2009[23]), and nearly one third of the University's total research awards come from the National Institutes of Health (NIH).

Key to understanding and responding to these trends is the recognition that scholarly production in the digital age requires the adoption of new information management practices to support the production, dissemination, and long-term stewardship of scholarly information. Just as the UC Libraries' strategic plans described above identified the aggregate challenge represented by these trends, other UC sectors and initiatives have reached similar conclusions. For example, in 2005 UC's Information Technology Guidance Committee declared that:

> *The University should create the capacity to manage scholarly digital assets in part by adopting strategies to ensure that the information produced in the course of research and instruction is effectively secured, managed, preserved and made available for appropriate use by others.* [24]

This perspective, and similar visions expressed both within and external to UC, constitute a call for the University and its libraries to devote energy and resources to the stewardship of the University's scholarly outputs as a core function. Such activities are a natural evolution of the library's traditional role to organize and provide access to global scholarship in support of research and teaching.[25]

The CDL's Role and Value Proposition

To meet evolving scholarly needs and extend the University's capacity to manage digital assets, the California Digital Library offers services to acquire, manage and preserve scholarly materials and to embed and integrate them into research, teaching, and learning. The CDL frames its work within the context described above by offering services that integrate the lifecycle of scholarly activities and the lifecycle of the results and products of scholarship.

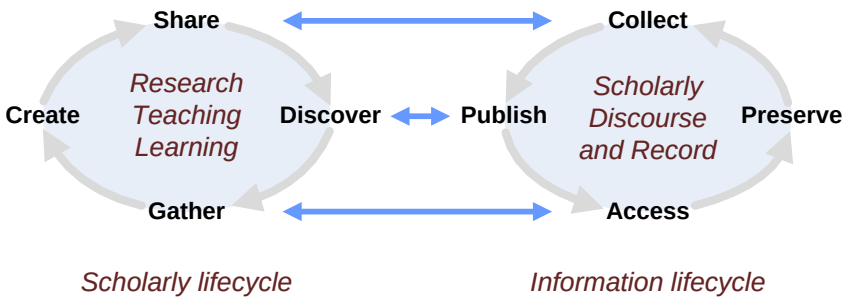

Figure 1: Interrelationship between the scholarly activity lifecycle and the information lifecycle.

Lifecycle framing is embedded in the CDL's mission and vision[26], and informs the CDL's work by:

- Ensuring that services are directly tied to UC's scholarly activities and needs;
- Enabling cost efficient service hand-offs and integration – anticipating prior and later stages in the life of a digital object or scholarly activity;
- Aligning services with activities in other sectors – identifying service offerings and partnerships that complement rather than duplicate efforts.

The CDL's lifecycle service offerings also provide value through the pursuit of economies of scope and scale, utilizing strategies such as centralization, modularization, and leveraging of information infrastructure components. These strategies seek to:

- Reduce otherwise duplicative costs and increase sustainability through consortial and single-point service models;
- Create extensible and scalable systemwide infrastructures – both social and technical - that leverage and highlight the unique strengths of individual UC campuses while supporting the common needs of all;
- Develop agile, interoperable services and strategies for information management that can be deployed across research, teaching, learning and administrative functions and integrated at appropriate levels;
- Cultivate strategic alliances within the academic community and beyond to amplify UC's reach and impact within the State of California and worldwide.

The CDL operates at a current annual cost-equivalent of $58 per UC faculty, student, and academic staff member.[27] When narrowed to a focus on the University's support of the 12,000 UC Senate Faculty members, CDL's services have a cost-equivalent of $1400 per FTE – one-third of one percent of the University's more than $400,000 in R&D expenditures per Senate faculty member.[28] Similarly, the CDL's portion of what the University spends on education per student ($15,820 on average) amounts to less than one half of one percent, or $72.[29]

Expenditure Per	CDL	University	CDL Cost as a % of University Expenditure
Student	$72	$15,820	0.46%
Faculty	$1,400	$414,289	0.34%

The return to the University on these investments has been summarized by former Director of Systemwide Library Planning Gary Lawrence (based on 2008 figures):

With a core budget of about $14 million, the CDL attracts an additional $18.5 million annually in voluntary co-investments from campus libraries and uses the resulting $32.5 million pool of funds to deliver about $52M in direct benefits to campuses, supports an additional $46M in measurable indirect benefits, and provides a technical platform and a leadership capability which fosters development of a host of service innovations that could not readily be supported by our ten campus libraries operating independently.[30]

CDL Programs and Services

The CDL's lifecycle services are offered through five complementary program areas:

Collect

CDL Collections develops and manages shared library collections that are available to users at all UC campuses. Working in close collaboration with campus libraries, the program manages UC's mass digitization efforts, oversees shared physical materials, and negotiates systemwide licenses for a wide array of digital content, cultivating strategic relationships with publishers to protect the University's interests (http://www.cdlib.org/services/collections/).

The value that collection services brings to UC's scholars and returns to the University includes:

- *World-class digital collections:* for a total annual expenditure that equates to $172 per student (equivalent to the cost of 1-2 print textbooks) and $3338 per faculty member (little more than the cost of a personal computer), UC students and faculty have access to a broad array of content critical to research and teaching, including 34,500 licensed online journals and 508,000 electronic books and databases — digital scholarly materials that are consulted and cited in UC's pace-leading awards of research grants and patents31 and help attract top scholars to the ranks of UC faculty — of digitized volumes from UC library collections (growing by nearly 1 million volumes per year). These costs include the fees paid to publishers and other content providers, supported through a combination of CDL funds and campus co-investment.
    - *A fraction of the cost of research:* CDL's staffing costs to develop and manage online content amount to less than 1/10th of 1% of research expenditure per faculty. [32]
    - *High use and low overhead:* Cost per use of these materials averages $0.81; only $.03 of this cost is attributable to the cost of staffing the service. The efficiency of CDL's licensing efforts is evident in comparison with other institutions: surveys and self-reported data from peer institutions reveal average per-use costs ranging from well over $1.00 per use up to $4.00 per use[33]
    - *Costs avoided:* $25 million annually through consortial license negotiations of 34,500 electronic journals and 500,000 electronic books and other databases (as compared to separate and multiple campus licenses).
        - *Value to campuses of systemwide licensing:* Another way to view the value of the CDL co-investments is through the prism of an individual campus. One campus library reports paying an average of $342 per title for non-CDL journal subscriptions that it licenses on its own[34], whereas that same campus's average cost per CDL-licensed journal title is $69. By this measure, every $1.00 spent on shared journal collections produces $4.00 in savings for the campus.

- *Digital transition at scale:* Mass digitization of legacy book collections is a complementary strategy that goes hand in hand with shared print collection management, while serving a broader purpose of adapting legacy collections to the needs of emerging digital scholarship. To date, more than 3 million UC books have been digitized in partnership with the Internet Archive and Google, at a rate of more than 800,000 annually. CDL manages the contracts and relationships with largescale digitization partners; funds, coordinates and sequences digitization operations at the campuses and regional library facilities; oversees technical aspects such as quality control and metadata; manages the integration of mass digitization outputs with other services such as OCLC WorldCat and HathiTrust; and exercises long-term stewardship over these systemwide assets on behalf of the campus libraries.
    - *Collaborative stewardship amplifies UC's collections:* CDL led UC's decision to join HathiTrust, a shared digital repository for mass-digitized and other content of which UC is a founding member. The HathiTrust corpus currently numbers over 7 million volumes from 36 partner libraries and is expected to grow to 12 million volumes or more by the end of 2014, rivaling Google Book Search in scope and breadth of content. Over 1.6 million volumes in HathiTrust (22% of the corpus) are in the public domain and thus fully accessible to users. UC's 1.8 million volumes ingested to date comprise 24% of HathiTrust today; a percentage that will grow as additional UC content is ingested into the repository.
    - *Economies of scale:* As both a preservation and an access service, HathiTrust offers significant economies of scale in managing and providing services across a large array of legacy content. Storage of UC's digital books in HathiTrust costs $0.12 per volume ($0.15 per volume when staffing costs are included).
    - *Enabling new modes of scholarship:* Text mining and other forms of sophisticated textual or computational analysis enabled by largescale digitization promise to unlock the knowledge contained in books in radically new ways. UC scholars are at the forefront of such efforts: seven UC scholars were awarded four out of 12 digital humanities grants awarded by Google in 2010 to explore use of the mass digitized corpus for digital humanities research. HathiTrust will be the locus of collaborative services to be built over the combined collections of its partner libraries, minimizing the cost to the partners of developing research-oriented applications.
    - *Reinvigorating print through on-demand services:* Print-on-demand (POD) services soon to be made available for UC's nearly 400,000 public domain volumes will offer convenient print copies at a reasonable cost to both the UC community and the general public. A POD contract recently signed with Hewlett-Packard Corporation will allow CDL to coordinate the distribution of UC books through multiple channels, including traditional retailers such as Amazon and on-site distribution at individual UC campuses. Revenue anticipated through these services will be allocated to offset the costs of managing the digitized books in consultation with the UC University Librarians.
- *Managing print collections more economically:* A significant challenge faced by the libraries in the inexorable transition from print to digital involves how to manage library investments in print collections. These concerns take two principal forms: the need to modify collecting practices prospectively to avoid duplicative print purchasing of materials that are increasingly less used in physical form, and the even larger corollary challenge of managing the vast legacy collections housed in UC's physical libraries, many of which are also duplicative and little used. *Shared Print* is a strategic initiative residing within the CDL Collections Program designed to facilitate and work in concert with the UC Libraries in the transition to digital information, allowing the libraries to reclaim and repurpose valuable campus space and ensure the breadth and long-term viability of UC's rich historical collections while paving the way for the redirection of capital and operational resources toward evolving digital concerns. CDL provides leadership to these efforts, including development

of service models, facilitation of extramural and vendor partnerships, and analytical and coordination support.

- The UC/JSTOR Paper Repository is a project undertaken in partnership with JSTOR in 2005 to create a page-validated dark archive of JSTOR journals built from UC print collections. The archive is built and managed at the UC Southern Regional Library Facility (SRLF) under the general direction and oversight of the CDL. To date, 1556 titles have been validated for the archive, comprising 55,443 volumes, over 200,000 issues, and 14.5 million pages. In 2009, Ithaka S+R released a decision support tool[35] to assist libraries around the country in making withdrawal decisions for JSTOR journals based on the presence of the dark archives maintained at Harvard and UC. Withdrawal projections developed by CDL in 2006 suggested that UC campuses can achieve over $5 million in capital ASF savings ($250K annualized) by de-accessioning local copies of the archived JSTOR titles. A limited amount of these savings have been realized to date.
- The Western Regional Storage Trust project (WEST) currently taking shape among sixty-plus libraries in the western region of the U.S. seeks to establish shared print journal archives across a broad geographic area, including program management and active ingest/validation of more than 260,000 volumes. WEST intends to recover, potentially, the space occupied by more than 1 million volumes now held in partner libraries, an aggregate savings of more than 200,000 linear shelf feet or more than 86,000 ASF reclaimed (equivalent to a mid-size ARL library). This effort, initiated and coordinated by the CDL, promises to transform the physical footprint of libraries throughout the region, and by extension, the nation, preparing the way for a more fully digital age.
- Additional efforts are underway to articulate a shared strategy for print monographs in light of the increasing availability of digital versions, such as those available via HathiTrust. In the short term, online access to digital surrogates promises to lower the cost of servicing UC's retrospective book collections by reducing the number of lending transactions and lessening the demand for physical delivery. In the longer term, shared strategies regionally and nationally may allow a managed draw-down of duplicative monographic collections both within UC and across a broader set of libraries.

Digital Special Collections (DSC) is at the center of collaboration between libraries, archives and museums both at UC and throughout the state of California to build integrated access to a world class digital collection of primary source material that serves from researchers and scholars as well as K-12 students.(http://www.cdlib.org/services/dsc/).

The value that digital special collections services brings to UC's scholars and returns to the University includes:

- *Mining California's gold:* the Online Archive of California (OAC) and Calisphere support scholars in their access to primary source material, with access to over 28,000 collection guides and 200,000 digital objects from 238 libraries, archives and museums throughout California at a cost of $4.27 per object and $0.51 per user session (unit costs decline as objects are added and usage increases). These collections represent unique and often rare historical evidence of interest to scholars worldwide. OAC and Calisphere attracted more than 2 million user sessions in the twelve months from October 2009 – September 2010.
- *Targeted for teaching*: built on the foundation of OAC, Calisphere supports the teaching of primary research methods at the K-12 and undergraduate level by making digital objects directly accessible, including 73 themed collections keyed to the California Content Standards. Calisphere themed

collection pages were viewed 3.4 million times between October 2009 and September 2010, at a cost per use of $0.09.

- *Cost avoidance:* DSC provides UC's libraries, archives and museums with technology, expertise, and best practices for building and providing broad, integrated access to their digital collections, thereby avoiding duplicative costs of building and maintaining such services locally. Some of this cost avoidance is in the form of systems to support archival collections management, such as the Archivist's Toolkit and Archon services currently provided by DSC. A typical commercial system of this type costs between $20,000 - $50,000 for a single campus.

Access

The CDL ***Discovery & Delivery (D&D)*** team connects UC faculty, students and staff to the content they need by integrating discovery into the environments where they work, while at the same time simplifying library systems. D&D provides reliable and seamless access to the University of California libraries' extensive research collections and beyond. (http://www.cdlib.org/services/d2d/).

The value that discovery and delivery services brings to UC's scholars and returns to the University includes:

- *Efficient discovery at UC-scale:* The Melvyl union catalog, one of the largest such catalogs in the country, provides a single place to discover all of the University's library collections, currently totaling 33 million items.
    - The current version of Melvyl costs $.31 per search, or about $.07 to process the more than 9 million transactions/year to keep Melvyl current. The Next Generation Melvyl initiative seeks to extend the scope of collections to global coverage (e.g., to almost 200 million items from 71,000 libraries worldwide) and move to a hosted service with the potential to reduce costs even further.
- *Last mile adds big value at little cost:* linking citations to content from 34,500 centrally-licensed electronic journals and 26,800 open access journals linked for online use through UC's link resolver technology, known as UC-eLinks .
    - Over 8 million uses of UC-eLinks services a year result in easy linkages between citations and the full text of articles or ebooks. At a cost of $.02 per transaction, the infrastructure for this essential last step adds only pennies to the cost of licensing the content.
- *Extending access to campus collections:* The Request resource sharing service manages 625,000 interlibrary loan requests annually with automated requesting and tracking functions. Roughly 70% of all borrowing and lending occurs within UC at one-third the cost of borrowing items outside of UC. (These costs are analyzed periodically by a systemwide Resource Sharing Committee.) CDL's investment of $.89 per transaction for the infrastructure that supports the system has the potential to become even more cost-effective based on recent analysis that could lead to streamlining.
- *Making descriptive data work harder:* The D&D team is developing systems to support emerging collection management needs for print and digital versions of books and journals. These activities build on deeply held expertise from thirty years of developing the Melvyl union catalog and using metadata to merge shared collection information.

Publish

The CDL ***Publishing Group*** provides open access digital publication services to the University of California academic community, supporting UC scholars and research units who seek unfettered digital

distribution of their research outputs in both traditional and non-traditional forms. The Publishing Group's transformative tools and platforms are designed to support new modes of scholarly communication and new communities of authors and readers. (http://www.cdlib.org/services/publishing/).

Though the early years of the CDL were, by design, characterized by a great deal of experimentation within the realm of scholarly publishing, in the past three years the CDL's publishing services program has directed its resources toward the focused design and development of a scalable systemwide publishing and dissemination infrastructure. This strategic repositioning of the CDL's publishing services springs from our recognition of the necessity of extensible and scalable infrastructure that positions the University of California to provide increasingly agile solutions for a host of timely concerns, including:

- Low cost publishing tools, for both well-established and emerging, grant-poor disciplines
- Branded publication management, for academic units, labs, Original Research Units (ORUs) and Multicampus Research Units (MRUs)
- Compliance with funding agency and federal mandates for access and preservation
- Assurance of peer review and authority within open access publishing
- Integrated access for publications that include data sets and other supplemental files
- Support for the dissemination of UC student research, both graduate and undergraduate
- Guaranteed online permanence

Providing open access publishing services supports the stated desires of UC faculty, on behalf of whom the UC Academic Senate has petitioned the University to work to advance open access initiatives – specifically naming the CDL among the strategic initiatives that should be supported.[36] By providing an aggregated open access platform for disseminating UC scholarly work, the CDL provides economies of scale that both increase the accessibility and impact of that work and lower the cost to the University as a whole of this dissemination.

The value that publishing services brings to UC's scholars and returns to the University includes:

- *Broad access at extremely low unit costs:* Nearly 300 UC academic departments and research units use CDL's *eScholarship* publishing platform to openly distribute their formal publications, technical reports, conference proceedings, seminar presentations, working papers, postprints, etc. The costs associated with this service compare extremely favorably to commercial publishing fees:
    - Average annual per-article cost for publishing a peer-reviewed article in *eScholarship*: $9.50
    - Typical per-article processing fee charged by publishers to make publication freely available in open access: $1500 - $3000
    - Average per-article publishing costs for subscription journals: $3800[37]
    - Average annual publishing cost for publishing an entire open access journal in eScholarship: $3260
- *Minimal cost to grow*: Because editorial costs are shouldered by participating units and publications and the platform technology is extensible, growth costs to add new contributors are currently negligible. The platform is also extensible to new types of publications (e.g., electronic theses and dissertations—in progress—and supplementary materials such as data sets that are becoming essential for scholarly dissemination).
- *Cost avoidance*: By providing a systemwide platform for the publication and dissemination of UC-sourced or supported research, the CDL eliminates the need for individual UC campuses,

departments and labs to develop and support similar infrastructure locally. This consortial model saves the system close to $500,000 per year in license fees alone. As the CDL moves toward an in-sourced solution for open access publishing and repository functions, further cost avoidance is achieved through the scaling up of services to serve additional communities and bodies of content (e.g., our current work on launching undergraduate and UCOP administrative repositories by July 2011 on the same platform, with minimal additional costs).

Preserve

The ***University of California Curation Center (UC3)*** offers a comprehensive suite of services to UC campus partners to support the long-term curation of the varied information resources underpinning the University's scholarly activities (http://www.cdlib.org/services/uc3/). Digital curation is the set of policies and practices focused on both maintaining and enhancing the value of digital resources through preservation and access for use now and re-use, often in ways unanticipated by the resources' original creators, in the future. In distinction to traditional paper-based library materials, digital resources are inherently fragile with respect to disruptive changes in technology and user expectation; without careful, pro-active curatorial management, digital information is easily susceptible to irretrievable damage and loss. Digital curation thus plays an important cultural and pedagogic role in mediating the transfer of knowledge across time and space, both within and across disciplines and generations of scholars.

The value that curation center activities brings to UC's scholars and returns to the University includes:

- *Securing the University's digital assets*. The information resources fundamental to the University's research, teaching, and learning mission have been acquired or generated at significant cost. Yet the value of that investment could be lost tomorrow if those resources are not prudently managed with timely monitoring and intervention to counteract the potentially destructive effects of technological obsolescence, accidental or malicious interruptions of IT services, or even something as seemingly harmless as "benign neglect." At a current annual cost-equivalent of $5.25 per academic faculty and staff the curation center operates a large-scale curation infrastructure, called Merritt, as a secure, controlled environment for the long-term curation of digital assets. The curation center's repository currently supports the archival management of over 1.6 million digital files with a total size of 79 terabytes, or TBs (79 trillion bytes, the digital equivalent of over 1.5 million physical books) at an annual cost of only $0.15 per file, of which 1/3 is recovered in storage fees from users. Since the total file size of most library collections is modest, for the annual cost equivalent of a safe deposit box (about $68), libraries can ensure the long-term usability of an average-sized collection of digital resources in Merritt.
    - In operating Merritt, the center employs accepted community best practice strategies of wholesale replication of digital content in disparate locations and heterogeneous storage technologies: constant monitoring to detect and repair bit-level damage; technology watch for recognition and response to incipient technological obsolescence; and disaster recovery and business continuity planning. The 93 institutions currently using Merritt include seven UC campuses, UCTV, three CSU campuses, the CDL Digital Special Collections and Publishing programs, other universities and colleges, California public libraries, and local historical societies.
- *Overseeing cooperative solutions.* Not all digital content owned by the University needs to or should be managed locally. Commercially-published journals and books are a good example of 'commodity' content that is more appropriately managed at a network level, where the common needs of the scholarly community can be addressed with greater economic efficiency both to UC and to the

global information system as a whole. For example, Portico is used as the University's primary strategy for guaranteeing long-term preservation of licensed journal content, while HathiTrust is the cooperative solution of choice for mass digitized books. In each of these cases, the material is widely held by libraries, making a shared investment both more efficient and more valuable from a community perspective. Both services offer additional levels of service to the community that would be infeasible to duplicate on a local level: in the case of Portico, post-cancellation access across a journal's entire customer base; and in the case of the HathiTrust partnership, full-text searching and access across a broad corpus of aggregated content. UC3 carefully evaluates both the costs and the technological robustness of these services to ensure that they meet appropriate community standards and that costs and services are economical for the University. The average annual cost to UC for an ejournal title in Portico is $5.64, while an average digital book (54 MB) costs $0.15 in HathiTrust. Locally-created content, by contrast, tends to be less amenable to a network solution.[38] By advising on and overseeing both local and network-level preservation and curation strategies, the center ensures that the University's investments are managed using the most appropriate solution and that UC's internal resources are concentrated on uniquely held material where they can add the greatest value.

- *Information assets as living documents:* Scholarly discourse is inherently cumulative in nature, dependent on a foundation of previous research, analysis, and results. The addition of value to scholarly content can be enhanced by facilitating the creative use and novel repurposing of that content. The curation center has pioneered a set of curation micro-services[39] involving modular components that can be deployed independently over time to keep digital content viable. Examples include services to streamline the domain-specific description of curated content in support of persistent citation and efficient discovery, and mechanisms by which the consumers of information from the Merritt repository can annotate that content in light of their own knowledge and experience. There are many points in the lifecycle of a digital resource at which important decisions are made, often implicitly and without deliberation, that can have significant consequences for long-term curation efforts. Thus, the curation center provides best practice recommendations for curation and data management, and offers consultative services for University content creators and consumers to guide them in making favorable choices. The center's micro-services approach is already gaining significant recognition within the broader preservation and curation community.[40]
- *Building tomorrow's archives at scale:* The World Wide Web is fast replacing more traditional media as the 'first draft of history,' yet the persistence of this record threatens to disappear into a digital 'black hole' if cultural memory organizations fail to preserve it.[41] The Web Archiving Service (WAS) is CDL's response to this challenge, ensuring long-term access to what is now a dominant arena of public discourse and the primary publication stream for in-depth reports from federal, state, and local government agencies and non-profit research groups. WAS archives 19.6 TB of web-published research materials, including 4,316 sites captured, and 30,619 captures run. To provide a sense of scale, the current scope of WAS collections is equivalent to 363,630 digital books. The service is optimized for academic research and used by 18 institutions whose 86 curators are building 210 archives. Since the service went into production in November 2008 there has been a 19.5% average growth in capture activity each month by curators.[42] In comparison to a comparable service that costs approximately $3300 per collection,[43] the cost of WAS averages $1885 per archive currently with costs declining as usage grows.
    - WAS does more than just provide researchers with passive access to archived materials; it places the tools for building large scale data sets and collections of Web content directly in the hands of UC faculty and researchers as well as library curators. By centralizing the significant infrastructure and resources needed to conduct large-scale Web archiving, CDL delivers research tools to faculty that would otherwise be inaccessible.

III. Prioritization and Decision-making

The CDL has a governance structure (see Appendix 1) and decision-making processes that ensure cost-effective alignment with the mission of the University and the constituencies it serves. We maintain a documented set of organizational goals and objectives (see link in box below) with a four-year horizon. Goals and objectives are driven by a vision of elevating the digital library for UC so that it becomes "expansively global and deeply local". The focus is on supporting access to the highest quality scholarship, providing a full range of services to support digital formats (including UC's unique digital assets) throughout their lifecycle, integrating access and services within our users' research and teaching activities, maximizing UC's effectiveness by offering significant economies of scope and scale, and using partnerships and alliances to elevate services to the network level for global reach and impact. Goals and objectives developed in the service of this vision benefit from regular consultations with the UC campuses and our collaborators (including, for example, structured interviews and full-day site visits at each UC campus as recently as Fall 2009).

CDL's 2009-2012 Goals

Support the UC scholarly research and teaching lifecycle

Expand researchers' access to UC's digital resources worldwide

Provide trusted and enduring access to California's cultural and intellectual heritage

Advance the state of the art in digital library services

Extend efficiencies and effectiveness

[Goals and supporting objectives can be examined in detail at http://www.cdlib.org/about/mission.html]

We set, adjust, and prioritize our activities by answering a set of key questions:

- Does the initiative, decision, or project support the UC mission?
- Is it responding to a clearly demonstrated need? Does it anticipate and support user needs? (Data driven, evidence-based)
- Is the need systemwide? Is there an opportunity to create efficiencies of scope and scale or to avoid redundant or duplicative costs undertaken by multiple campuses?
- Does it have a well articulated sustainability model? Is it scalable? Extensible? Repurposable? (by ourselves, our UC partners, the larger digital library community)
- Can we leverage a solution that already exists? On a campus? Outside of UC?
- Does or will it lower costs for the UC system?
- What opportunity costs does it represent — what current or next objective will need to be adjusted or foregone?
- Measured against the above criteria, has the service outlived its usefulness?

Below are several *decision profiles* which highlight the results of the CDL's prioritization methods.

1. *Next Generation Melvyl—leveraging a network solution to serve systemwide needs.* In 2005, the UC libraries' Systemwide Operations and Planning Advisory Group (SOPAG) charged a group to inventory current bibliographic practices that had grown up organically, and to develop a compelling future vision to motivate change within the libraries. From that charge came the widely-discussed *Bibliographic Services Task Force report,*[44] key among whose recommendations were to "Create a single catalog interface for all of UC" that supported "searching across the entire bibliographic information space."[45] After reviewing the recommendations, the University Librarians chose to enter into a pilot project with OCLC to evaluate WorldCat Local as a replacement for UC's current union catalog, Melvyl. User assessment studies have confirmed

the value of providing access to all of the world's content, while linking to UC's local delivery services.

2. *DataONE—a clearly-demonstrated need central to the UC mission: long-term access to scientific data*. Sustainable long-term access to scientific data is crucial to the UC research enterprise. The National Science Foundation's DataNet solicitation recognized that key to the success of this effort was the involvement of libraries.[46] On behalf of UC libraries, UC3 has engaged as an active member of DataONE,[47] which is developing the cyberinfrastructure and organization to support the preservation of and access to observational biological and environmental data critical to the study of climate change, an area of significant research activity at multiple UC campuses. By working collaboratively with UCSD, UCD, the UCSB Library and National Center for Ecological Analysis and Synthesis, and a range of other partners on this large scale international initiative, UC3 is working to help chart the future role of libraries in long-term data curation and ensure that services are appropriate and available to the UC community.
3. *DataCite—anticipating user needs in response to research dissemination trends.* Creating a sustainable cyberinfrastructure for referencing and sharing of data sets has emerged close to the top of the international research agenda. The UC community produces an enormous body of data that is both inaccessible and at-risk, and a pragmatic first step towards curating it is to establish a way to permanently register and record the existence of each data set. To that end UC3 is a founding member of DataCite,[48] a global consortium of academic and scientific memory organizations designed to create data publishing opportunities for producers. Rather than each campus investing independently in DataCite, UC3's intellectual, technical, and organizational investment allows the UC community to share, provide access to, and to get credit for research data.
4. *Value-based Journal Licensing —evidence-based negotiation designed to reduce costs.* CDL is leading a refinement of the UC Libraries' analytical approach to purchasing and licensing scholarly journals through a data-driven methodology that uses key vectors of value including utility, quality, and cost effectiveness to evaluate individual journals contained in large publisher packages and assess the value of the package as a whole. By supplementing human expertise and judgment with objective measures in selecting essential materials to support UC's research and scholarship, CDL is able to achieve significant cost savings for the system in support of the University Librarians' goal to realize 15% savings in licensing costs in FY2009-2010 and FY2010-2011.
5. *CDL Hosted Archivists' Toolkit/Archon Service—a community sourced solution that avoids local costs:* The California Digital Library hosts the open-source archival data management systems Archivists' Toolkit (AT) and Archon for use by contributors to the Online Archive of California. AT and Archon are complementary systems that provide integrated support for records and manuscript collections. In April 2009, CDL conducted a lightweight needs assessment of OAC contributors to gauge interest in a potential hosting service. The survey showed widespread interest in a centrally managed service (36 institutions out of 55 respondents, or 71% of the total). After follow-up conversations with individual contributors and a review of the cost/benefit to CDL, DSC decided to mount a pilot service beginning in March 2010. Support for AT and Archon furthers CDL's mission to support digital access to archival information. Collection guides and encoded digital objects generated by these tools can be contributed to the OAC, streamlining both creation and ingest and offering partner institutions an efficient method for contributing content to CDL services without the need for local infrastructure. Existing staffing was able to be leveraged to implement the databases and provide ongoing support for new contributor workflows; the state-of-the-art software is open source and thus entails no license fees; and ongoing maintenance requires minimal

overhead. As well, AT was originally developed at a UC campus (UC San Diego), where expertise is available to assist with systemwide adoption. Finally, a Mellon-funded planning project to merge AT and Archon into a single product called ArchivesSpace is underway with active UC involvement; by supporting both products now, CDL is well-positioned to provide a seamless transition to the new product for its contributors and to work with the project's sponsors on a hosted services sustainability model.

6. *eScholarship - leveraging the combined strengths of CDL and UCPress to support community needs*. A joint service of the CDL and UC Press, UCPubS provides a suite of open access digital and print publishing tools to UC centers, institutes, and departments that produce scholarly books. UCPubS was established in response to the substantial need for publishing infrastructure support with the UC system, as identified through surveys and meetings with Vice Chancellors for Research, Deans and faculty about their current publishing and dissemination practices. A year-long business planning process revealed opportunities for the CDL and UC Press to address these systemwide needs by establishing an integrated open access publishing, print-on-demand, distribution and marketing program. Identifying the market early on and tailoring the services to clearly-defined needs has paid substantial dividends: UCPubS participation has grown enormously in the two years since its launch, with over 60 titles now enrolled in the program representing a burgeoning new revenue source in support of UC's publishing activities. The CDL and UC Press are now working to design a broader publishing services program to extend this successful collaborative model to other promising content areas, including journals and grey literature. This program is part of the University of California's broader effort to ensure a sustainable scholarly publishing system in support of the University's research and teaching enterprise.
7. *Metasearch—evidence-based decision-making.* An ongoing challenge for libraries is how to integrate access to the multitude of specialized indexes, portals and other licensed resources. In 2008, CDL and UCLA entered into a pilot project to assess metasearch functionality, a technology that federates searches across multiple resources. Based upon Ex Libris's Metalib product, CDL configured a specialized portal for women's studies and UCLA conducted the user assessment, informed by goals from the Heads of Public Services since other libraries were interested in a similar solution. The premise of the experiment was that CDL could host the infrastructure for campus libraries to create individually or collaboratively various subject aggregations of specialized resources. The assessment results confirmed that users valued the concept of a specialized portal but the promise of the technology did not outweigh its challenges. CDL incorporated the assessment results into a cost/benefit analysis and made the decision not to implement the software.
8. *Counting California—knowing when to retire a service*. Counting California (CoCa) was an award winning, innovative site, created by the CDL in 2001 to provide integrated access to a range of California statistical information from state, federal and local agencies. Although CoCa was among the earliest sites providing aggregated statistics on the web, a 2008 analysis showed that it no longer met a key need. An end user survey revealed that most users sought population information from newer sites that had become widely available on the web, and most of these sites provided equivalent content with a richer feature set that would have been costly to develop. As a result the service was retired, users were directed elsewhere, and resources were redeployed to projects that better served UC priorities.
9. *Digital Preservation Program (now UC Curation Center)—efficiencies of scope and scale applied to a new challenge*. In 2001 the UC University Librarians recognized that the long-term retention of digital materials was an urgent problem for the UC Libraries. They also recognized the potential to work collaboratively to address the problem. Rather than each campus individually

building the significant organizational, technical, and intellectual infrastructure they established the Digital Preservation Program (now UC Curation Center) in 2002 at the California Digital Library. This collaborative model enables the UC libraries and the campus community to cost-effectively act as guardians over UC's digital scholarly assets without having to individually invest in the requisite deep technical expertise and infrastructure. After several years of operating the Digital Preservation Repository, CDL determined that campus users found it difficult to use, and the cost/benefit ratio had not kept pace with more recent developments in storage and preservation technology, causing a retooling of the entire approach. As the UC Curation Center has matured, so have the resources and savings to campus community. With the launch of the Merritt and the suite of curation micro-services, UC campuses will have access to flexible tools to meet local needs as well as a next generation centrally-hosted preservation solution that is now more cost-effective to maintain.

IV. Sustainability

Libraries historically have been seen as core infrastructure, i.e. as cost rather than revenue centers—but from its founding the CDL was set within a context of sustainability planning, charged to contribute both to the sustainability of UC library services and to that of the entire enterprise for managing scholarly information.[49] Today, that effort has assumed renewed importance.

A recent Ithaka report defines sustainability as "*the ability to generate or gain access to the resources – financial or otherwise – needed to protect and increase the value of the content or service for those who use it*."[50] The Ithaka Report focuses on five cultural and structural aspects of sustainability:

1. Dedicated and entrepreneurial leadership – comprising R&D, and an openness to experimental approaches and the value of experimentation
2. A clear value proposition
3. Minimizing direct costs – in-kind support, outsourcing, partnerships/collaboration, volunteers
4. Developing diverse revenue streams – revenue, grants, subsidies, hybrids
5. Clear accountability and metrics for success

CDL has recently gauged its efforts in light of Ithaka's analysis. While attention to sustainability is embedded into our decision-making processes, not every CDL service is or can be economically self-sustaining. CDL will continue to rely on UC core funding for some of its legacy, developing, or aspirational services, realizing at the same time that some services may need to increase their base of external support. Nevertheless, when viewed as a suite or portfolio of academic information services, the CDL employs a solid business modeling approach that does, in fact, minimize costs, develop revenue, is accountable and metric-driven, and is entrepreneurial.

The CDL's approach to Ithaka's second factor – a clear value proposition – is described above in a dedicated section of this document. Evidence of the degree to which sustainability is structurally and culturally embedded in the CDL along the remaining four dimensions identified by Ithaka includes:

Entrepreneurial approach: CDL consciously cultivates a research and development mindset and capacity, proactively scans its environment, promotes creativity and appropriate risk-taking, and develops outward-facing leadership, by:

- Explicitly declaring that innovation, collaboration, sustainability, and openness are part of its vision, values, and strategic planning (http://www.cdlib.org/about/mission.html).
- Embedding research and development processes and capacities into its planning and operations – "discussion forums" and "tech talks" are scheduled bi-weekly and monthly respectively; "explore teams" are assembled for project initiation; a separate IT development environment is maintained.
- Recruiting leaders – program directors, project managers, and technical leads – with diverse experience and perspectives. Current staff leaders have previous experience at Microsoft; Harvard; UC San Diego; Brobeck, Phleger & Harrison LLP; UC Berkeley; Interwoven; TRW; and a variety of other non-UC public and private sector organizations.
- Recruiting graduate student interns from local graduate programs, who bring with them fresh skills and perspectives – an average of two interns per academic semester over the last several years.
- Hosting national and international conferences and visitors – ICOLC 2008 and iPres2009 were both hosted by CDL; in the past year delegations have visited from Japan, China, the EU, and Bahrain.
- Assembling diverse partnerships and partners – as described in various sections of this document.

Minimizing direct costs: CDL minimizes costs to the University as a whole as well as controlling its own direct costs.

- CDL ***minimizes the University's costs*** to manage scholarly information through economies of scope and scale by developing shared or reusable technical tools and capacities, organizational capacities, and policies and best practices. In this regard the very existence of the CDL is a sustainability strategy itself, through which a range of activities and infrastructure are single-sourced instead of duplicated at each UC campus. Specifically, the CDL:
    - Licenses, builds, and hosts tools that support efficient resource sharing among the UC campuses. Our tools for interlibary loan (*Request service*), online linking (*UC-eLinks*), and online discovery (*Melvyl*), as well as the Shared Cataloging Program operated at UC San Diego and managed by the CDL, allow the UC libraries' collections to be used as a single systemwide collection, reducing the need for each campus to make separate purchases. Uses campus "co-investment" funds to license digital content on behalf of all UC libraries and campuses, creating a massive shared collection that features efficient negotiations with vendors, discounts for volume purchasing, and management economies.
    - Builds curation tools (the shared *Merritt Digital Preservation Repository* and UC Curation Center's *Micro-Services*) that are designed for easy adoption and use, lowering costs for UC units to meet preservation and related curation needs.
    - Builds publishing tools separately and in collaboration with UC Press (the *eScholarship* platform and *UCPubs* service suite) to support the dissemination of UC's scholarly output in all forms, avoiding separate campus or departmental investments in local solutions.
    - Manages partnerships at network scale, for example by representing all UC campuses in a partnership with Google to digitize books, and enabling the use of those millions of digital books alongside other digitized collections from around the U.S. by representing UC within *HathiTrust*.
    - Controls subscription costs: In addition to the significant inherent savings resulting from the consortial licensing of content, CDL employs specific methodologies to assess value and determine fair prices for subscription journals with sustainability and cost control in mind. For example, journals from commercial publishers are benchmarked against non-profit journals in the same discipline using the Bergstrom-McAfee Journal Prices index; these data are used in negotiation and decision-making in an attempt to counteract the high profit margins sought by some publishers.

- CDL ***minimizes internal operating costs*** through a number of efficiency strategies that include:
    - *Zero-based budgeting* which allows annual (essentially continual) adjustment of resource allocations.
    - *Identifying common programmatic needs* and organizing to provide them with utility internal service groups such as infrastructure, assessment, user services, project management, and business services.
    - *Securing campus co-investments of expertise and other resources* instead of assuming all direct costs for development or management of initiatives and services. Examples include:
        - UC-eLinks – powered by software hosted at the CDL but co-purchased and co-maintained by the campus libraries;
        - eScholarship document vetting, deposit and metadata enhancement – performed by campus-based "site administrators" assigned by each participating unit;
        - Licensed content – quality control and first-level vendor relations performed by campus-based "resource liaisons" coordinated by CDL;
        - UC3 – engages teamwork and technical input from UCLA, UCSD, UCB, UCM, the San Diego Supercomputer Center, and the Texas Advanced Computing Center, in exchange for access to UC3 analyses and tools.
        - Service specification and assessment – constructed through multi-campus committees and campus-based sessions.
    - *Leveraging external partnerships with a focus on collaboration*. Acknowledging that there are organizational and financial limits on CDL capacity or expertise, we build partnerships and manage collaborations to distribute tasks where expertise, capacity, and interest or shared-goals are located. This has the additional and equally valuable effect of increasing the impact of our initiatives due to more and more varied stakeholders, contributors, and, often, users. Examples include:
        - HathiTrust – as a founding partner, the CDL avoided the costs of creating a local preservation and access repository for its digitized books, increasing the value of those books by placing them in context with millions more from other R1 university collections and joining forces with the combined talent pool of the HathiTrust partners to develop research services;
        - Web Archiving Service – the principal features, workflow, and user Interface were specified and assessed by a multi-institutional and multi-sector team.
        - UC3 – builds on specialized technical, metadata and preservation expertise from a range of stakeholders with linkage to national and international efforts and research communities such as DataOne;
        - DataCite – as a founding member of this international consortium (with its own sustainability plan), the CDL contributes to and draws from resulting standards for exposing and linking datasets to other scholarly records, a core scholarly information lifecycle challenge.
    - *Engaging unpaid graduate student internships* for project specific work. Note that this also helps establish skills and knowledge among the next generation of information professionals, including those from UC Berkeley, and is an outreach activity. Examples from the 12 interns in the last three years include:
        - Analysis of the state of mobile technology used for library services; public relations/marketing project for mass digitization program (2010)
        - Planning for the international iPres Annual Conference (2009)

        - Automating subject classification of eScholarship documents (2008)
    - *Outsourcing where it makes sense, insourcing where it doesn't*. CDL analyzes and pursues those cases where services, software, or service components can be outsourced more inexpensively than buying or building for in-house operation. In general, outsourced solutions work best for achieving economies of scale, where services are mature or have become commodities, or for areas that are not related to CDL's core competencies. Examples include:
        - IT infrastructure – our development and staging infrastructure was consolidated and moved to the UC Berkeley data center when it became inefficient and expensive to manage on-site. Our end-user and internal problem reporting and ticketing system is a hosted solution at approximately 30% of the cost (in FTE equivalence) of internal operation. Multi-terabyte storage for UC3 is being transitioned from local to SDSC and Texas Advanced Computing Center storage where scale yields cost efficiencies compared to local management and procurement. Not only are economies of scale greater for these high volume computing centers, but they also provide replication solutions to ensure business continuity and standard preservation practice.
        - Mass digitization – CDL-brokered arrangements with Internet Archive and Google have produce digital copies of millions of UC books and other items with minimal direct costs (and zero equipment costs). In calendar year 2009, the mass digitization group supported the digitization of 836,000 volumes at a cost of $1 per volume. Funding this digitization directly rather than in partnership with external sponsors would have cost the University $27.5 million dollars.[51]
        - Next Generation Melvyl – Currently, the cost of running Ex Libris Aleph software in maintenance mode totals $610,000 for minimal staff, infrastructure, software licenses and MARC records. By moving to a hosted model, it is likely that a savings of approximately $245,000 could be realized.
        - eScholarship – the eScholarship service currently runs on an outsourced platform available from bepress. CDL is re-architecting the service in FY2010-11 to use UC3's Merritt repository instead, eliminating the cost of this license (currently 25% of the eScholarship budget), increasing its ability to develop new services, and ensuring preservation-level archiving of eScholarship content. In this instance, an in-sourced solution will provide better value to CDL and its contributors.
    - *Leveraging open source tools and ethos*. We use open source tools and contribute to open source projects for several reasons, including as a cost avoidance strategy. Other benefits include amplifying costly development resources through community contributions and expertise. Examples include:
        - Web Archiving Service – uses open source web crawling tools (heretrix)
        - Micro-services, BagIt, JHOVE – the CDL is leading the development of these curation tools but has also partnered with other institutions and benefited from grant funding to support their development.
        - Xtensible Text Framework[52] – commissioned by the California Digital Library (CDL) to be the primary access tool for its collections, the eXtensible Text Framework (XTF) provides a powerful, flexible platform for providing access to digital content. It consists of Java and XSLT 2.0 code that indexes, queries, and displays digital objects. While the software is actively maintained and supported by CDL developers, its conversion to an open source project also allows a large group of adopters and adapters to return improvements to the CDL at no cost.

        - Archivists Toolkit/Archon – the CDL supports these open source tools as the recommended set for UC libraries and museums and other partners who contribute digital special collections materials to the Online Archive of California and Calisphere, but incurs close to zero cost in doing so.

Developing diverse revenue streams: CDL has begun to explore new avenues for diversifying revenue from external sources. While many of these plans are still aspirational, others have begun to bring in revenue with projections for the future.

- **Direct revenue:** As a means of supplementing ongoing operational costs, CDL is exploring revenue generation in several service areas.
    - Web Archiving Service (WAS): With the assistance of Ithaka, CDL developed a business plan aimed at determining reasonable fees for UC users and a plan for external users that would generate service fees to cover the marginal costs of storage plus surplus to defray some of the service's ongoing expenses. This plan has been launched and has attracted five external users, on target for revenue projections.
    - Print-on-demand: a service agreement with Hewlett Packard will provide print-on-demand for UC's out of copyright digital books. While current estimates are speculative, comparison with the results at the small number of peer institutions that have implemented similar programs (Michigan, Cornell) suggests a revenue potential of several hundred thousand dollars per year. Surplus revenues will be used to offset any direct costs of providing the service (including CDL administrative costs), underwrite the cost of HathiTrust participation, and other uses as approved by the University Librarians.
- **Multiple approaches:** As new services are developed, especially those with audiences beyond UC, CDL is consciously including sustainability planning in the early phases. A good example of this is EZID, CDL's identifier service which includes DataCite DOIs and another identifier scheme called ARKs. Our goal for EZID will be cost recovery, and to reach this goal, we will pursue a diverse mix of solutions. These may include use fees, grant-supported development, partner cost-sharing, and more.
- **Publication costs:** To address the unsustainable structural problem resulting from increasing costs to license scholarly journals and the downward trend in library budgets, CDL is analyzing and investigating means of diversifying funding for journal publication, including the role of author fees, submission fees (both of which might be covered by grant funding), and differential pricing for value added services.
- **Donations, gifts, and sponsorship:** Although CDL does not cultivate donors in the same way that campus libraries do, there are increasing opportunities to use this approach for both small and large cost subsidies. This is an area we plan to explore further.
    - Donation of hardware and storage: Given that storage and hardware costs can be significant for areas such as digital curation and preservation, CDL has been exploring donations from vendors and other providers in exchange for sponsorship credit for commercial vendors and/or research and development opportunities. For example, the Texas Advanced Computing Center, University of Texas has agreed to provide storage at no cost on a trial basis in order to explore the needs of a range of customers.
    - The Moore Foundation made a gift to CDL to further explorations of data curation.
    - In-kind donations of tools to support collaboration and project management have aided in experimentation with different methods without incurring upfront costs.
    - As a means of ensuring that future costs for digital curation can be met, CDL and some of its partners are investigating the feasibility of endowing resources such as a fixed amount of storage, the digital equivalent of endowing a physical library space.

- **Grant funding:** Since its inception, CDL has generated an average of $1.2 million a year in grants from a range of government and private sources[53] and has received $1.9 million in 2010-2011 alone with several other grants pending.
    - Nearly one-third of the Digital Special Collections program budget in 2010-2011 is underwritten by grant income to support CDL's development of new tools for content contributors and end users. Additional grant funds support the processing of new archival collections among DSC's partners (both within and beyond UC), expanding access to those collections within OAC.
    - UC Curation Center: Since the inception of the digital preservation program, it has attracted an average of $1 million a year from a diversity of funding sources including the Library of Congress National Digital Information Infrastructure and Preservation Program, the National Science Foundation, the Library Services and Technology Act, the Moore Foundation, and (pending), the National Institutes of Health and Microsoft Research.
    - Grants are especially effective for funding new development costs and experimental ventures but are not suited for funding operational services. All grants require or we insert a sustainability component for post-grant support to move from project to program. For example,
        - The Web Archiving Service received funding specifically for a consultant to develop a business plan.
        - Local History Digital Resources Project – an IMLS-funded program administered by the California State Library that has brought over $1 million in grant funding to CDL over the past ten years, supporting digitization of thousands of historical materials at libraries and archives throughout the state. Through successive grant cycles, CDL has worked with other California organizations to build technical capacity and expertise among the many institutions that have been included in the program since its inception and has developed technical standards and procedures to streamline operations. Ongoing operation of the program has now been transferred to another cooperative library service agency in California, allowing CDL to concentrate on more advanced development projects.

    Another benefit of obtaining grants is developing partnerships, many of which extend beyond the grant period. Collaborations with Stanford, University of Virginia, New York University, University of North Texas, UC Berkeley (with different campus units including the Information School), the Library of Congress and others have led not only to multiple grants but to productive relationships that bring new technologies and professional development opportunities to staff.

Clear accountability and metrics for success: Accountability begins by setting goals and objectives and is followed by mechanisms to measure progress and to revisit and rebalance as necessary.

- **Goals and objectives** – CDL's goals and objectives for 2009-2010 are based on input from advisory groups, interviews with campus libraries, evidence from user studies, and awareness of environmental trends. Objectives for 2010-2011 were updated with course corrections as needed.
- **User studies** – CDL's User Experience Design Service plays a part in almost every service at some point in its development. A recent investigation of mobile services[54] surveyed user habits and preferences, campus library mobile services and aspirations, and vendor offerings to recommend how CDL services could add value. At the other end of the cycle, usability studies and usage analytics have contributed to decisions to discontinue services. This approach is deeply embedded in CDL's culture and practice with several staff members who are trained in assessment techniques.

- **Advisory structure** – CDL participates in the UC libraries' consultative structure [see Appendix 2] at every level but also seeks advice from its own Digital Library Services Advisory Committee and Joint Steering Committee for Shared Collections. Within the Office of the President, the Systemwide Library and Scholarly Information Advisory Committee (SLASIAC) advises the Provost on matters that influence CDL's work, and CDL works closely with the Academic Senate's University Committee on Library and Scholarly Communication (UCOLASC). CDL regularly reports to and seeks advice from all of these groups.
- **Metrics and Key Performance Indicators** – Recently, CDL programs have been developing Key Performance Indicators to delve more deeply into ways to monitor trends and determine value for services. As a first step, most services contribute analytics to the Executive Director's "dashboard" that is updated regularly to show trends in size of content and usage. Examples of specific uses include:
    - Request service component: VDX is the vendor software used to track interlibrary requests within UC. Key Performance Indicators revealed it is an expensive service due to the complexities of the process and customization for local practices. It is likely that efficiencies can be found both for CDL and campuses by rationalizing the update process and providing additional training for campus staff. Further efficiencies will require a concerted effort to examine and streamline the entire process, but the KPIs flagged the service as one to analyze more carefully.
    - eScholarship: CDL staff recently analyzed service costs and trends for consideration in merging services with UC Press. This analysis was also used to compare costs against other open access alternatives as well as commercial publishers.
    - Licensing: CDL regularly benchmarks journal licensing performance via the several value-based algorithms described earlier, and also reports regularly on its licensing activities, including savings and cost avoidances, to both the University Librarians and UCOLASC.

Conclusion

The pace of change affecting libraries is unlikely to lessen, whether driven by financial exigencies, penetration of new technologies, or the evolution of scholarly practice. CDL seeks to remain flexible and responsive but also anticipatory by following its vision to make the digital library "expansively global and deeply local:" global through the provision of access to research collections worldwide, open dissemination of UC scholarship where feasible, and strategic alliances at state, regional, national, and international levels, and deeply local by ensuring that UC-sourced information assets and unique collections are accessible and manageable now and for the future.

To meet these challenges, the CDL must make strategic choices that benefit the UC community and support the University's mission as a public institution of higher education. Guided by a compelling value proposition, appropriate methods for prioritization and decision-making, and an approach to sustainability that can serve the CDL in the years ahead, we are confident that the University and the CDL can marshal the resources necessary to sustain the scholarly research practices and output of the University of California for present and future generations of UC scholars.

Appendices

1. Organization chart: http://www.cdlib.org/about/docs/CDL_Org_Chart.pdf
2. UC Libraries Consultative Structure: http://www.cdlib.org/groups/docs/UCL_Consult_Structure.pdf

---

[1] For example, "Information is the lifeblood of colleges and universities." In Bess, James L. and Jay R. Dee. Understanding College and University Organization. Vol. II Dynamics of the System. Stylus Publishing. 2008. P. 643.
[2] Henry, Charles. "The Idea of Order" in The Idea of Order: Transforming Research Collections for 21st Century Scholarship. Council on Library and Information Resources. June 2010.
[3] Library Planning and Action Initiative Advisory Task Force Final Report. University of California. March, 1998 p.6.(http://www.slp.ucop.edu/lpai_new/finalrpt/index.html)
[4] See "Strategic Directions for Libraries and Scholarly Information: Universitywide Planning & Action" at http://libraries.universityofcalifornia.edu/planning/strategic_directions.html
[5] The California Digital Library: 2008 Profile, p. 3. (http://www.cdlib.org/news/docs/CDL_Profile_2008.pdf)
[6] For a list of services, see http://www.cdlib.org/services/
[7] As indicated, for example, by grants (http://www.cdlib.org/cdlinfo/category/grants/) and awards (http://www.cdlib.org/cdlinfo/2010/05/19/patricia-cruse-digital-preservation-pioneer/; http://www.cdlib.org/cdlinfo/2005/09/08/oac-receives-the-society-of-american-archivists-coker-award/)
[8] http://ucfuture.universityofcalifornia.edu/
[9] UC Commission on the Future. Second Round Recommendations from Working Groups. June 2010. P. 64 (http://ucfuture.universityofcalifornia.edu/presentations/cotf_second_recs.pdf).
[10] *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue Ribbon Advisory Panel on Cyberinfrastructure*. 2003. http://www.nsf.gov/od/oci/reports/atkins.pdf and *Our Cultural Commonwealth: The report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*. 2006. http://www.acls.org/cyberinfrastructure/ourculturalcommonwealth.pdf
[11] "Higher-performance *computation* provides more powerful tools for discovery through analysis and more systemic and realistic simulations. Acquisition, curation, and ready access to vast and varied types of digital *content* provide the raw ingredients for discovery and dissemination of knowledge. Computation and content, integrated through networking, offer new modes of *interaction* among people, information, computational-based tools/services, and instruments." Ibid, p. 44. (Highlighting in the original).
[12] op cit., p. 7.
[13] UC San Diego's Software Studies Initiative's "Visualizing Patterns in Databases of Cultural Images and Video." is a recipient.
[14] UCI and UCLA are among the six founding members of BIRN (http://www.birncommunity.org/about/overview/)
[15] See University of California Library Statistics, available at http://libraries.universityofcalifornia.edu/planning/stats/index.html.
[16] UC Office of Scholarly Communication. *Postprint Repository Services: Context and Feasibility at the University of California*. 2005.
[17] http://ucsdnews.ucsd.edu/newsrel/general/03-0821stCenturyPeerReview.asp .
[18] See, for example, the Electronic Cultural Atlas Initiative, based at UC Berkeley (http://www.ecai.org/).
[19] At this writing the Registry of Open Access Repository Material Archiving Policies listed 46 funders whose research grants included requirements for public access, among them NIH, the Wellcome Trust, and the Howard Hughes Medical Institute. Pending U.S. legislation – the Federal Research Public Access Act – would effectively extend the requirement to all significant federal agency funders.
[20] Faculty are growing resistant to the notion that publishers hold exclusive rights to their research. Publishers are looking for opportunities to create new kinds of commercial content from new modes of scholarly communication. Libraries continue to balk at unsustainable license fee increases that stretch budgets beyond feasibility.
[21] http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928&org=NSF. For an example of how CDL services are supporting this policy, see this blog entry at the UC Davis Health Sciences Library, September 30, 2010. http://blogs.lib.ucdavis.edu/hsl/2010/09/30/the-nsf-and-the-dmp-or-the-national-science-foundations-soon-to-be-implemented-data-management-plan-how-it-will-affect-you-where-you-can-turn-for-help/

[22] NIH Data Sharing Policy. http://grants.nih.gov/grants/policy/data_sharing/
[23] National Science Foundation Award Summary: By State Institution FY2009 (http://dellweb.bfa.nsf.gov/AwdLst2/default.asp)
24 "The University should create the capacity to manage scholarly digital assets in part by adopting strategies to ensure that the information produced in the course of research and instruction is effectively secured, managed, preserved and made available for appropriate use by others." from Creating a UC Cyberinfrastructure: Report of the University of California Information Technology Guidance Committee. December 2005. http://www.universityofcalifornia.edu/itgc/ITGC_final%20report.pdf
[25] For examples: a) the UC Libraries' Collection Development Committee declared that "UC faculty and researchers in collaboration with UC Libraries are increasingly engaged as creators of digital content that requires ongoing management and long-term stewardship." In The University of California Library Collection: Content for the 21st Century and Beyond. July 2009. p.1; b) Abby Smith states that "The library is likely to provide repository infrastructure for stewardship of university-based information assets." In No Brief Candle: Reconceiving Research Libraries for the 21st Century. CLIR, August 2008. P. 18.
[26] http://www.cdlib.org/about/mission.html
[27] CDL's base budget of $16.8M divided by 289,000 which represents UC's 232,000 students, and 57,000 academics (faculty and other academic staff) based on data from "Statistical Summary of Students and Staff: Fall 2009" (http://www.ucop.edu/ucophome/uwnews/stat/statsum/fall2009/statsumm2009.pdf .
[28] See UC Annual Accountability Report (May 2010), p. 134. http://www.universityofcalifornia.edu/accountability/documents/accountabilityreport10.pdf
[29] See The UC Budget Myths and Facts (January 2010) p. 4. http://www.universityofcalifornia.edu/budget/documents/myths_facts011110.pdf
[30] Lawrence, Gary. "The California Digital Library" in Business Planning for Digital Libraries: International Approaches. Mel Collier (ed.). Leuven University Press. 2010. [Note that recent service developments and reconfigurations prevent a repetition of Lawrence's analysis but we would expect the ratio of returns to be of similar magnitude.]
[31] A University of Illinois study recently concluded that three to six dollars is returned in grant awards in the sciences for every dollar invested in library collections ("Luther, Judy. University Investment in the Library, What's the Return. Library Connect. 2008.);
[32] For total research expenditures per Senate faculty member, see "UC Accountability Report", May 2010, p. 134 (http://www.universityofcalifornia.edu/accountability/documents/accountabilityreport10.pdf)
[33] See Franklin, Brinley, "Use Measures for Electronic Resources: Theory and Practice," 2005. (PowerPoint presentation) http://www.libqual.org/documents/admin/ALCTS_BRINLEY.ppt; OhioLink lib-stats mailing list posting, July 07, 2009 (http://www.lib-stats.org.uk/).
[34] Specifically, the library reported these per title averages of Science, technology, medicine titles: $520 (average cost of 1371 titles from EBSCO); Social sciences, humanities: $155 (average cost of 1146 titles from SWETS); All areas: $223 (average cost of 245 titles from Harrassowitz).
[35] See http://www.ithaka.org/ithaka-s-r/research/what-to-withdraw/print-collections-decision-support-tool.
[36] Academic Senate Letter to President Mark Yudof, June 16, 2009, available at http://www.universityofcalifornia.edu/senate/reports/MC_Yudof_open%20access%20FINAL.pdf. The University Committee on Library and Scholarly Communication letter of May 14, 2009, on behalf of which the aforementioned Senate letter is a transmittal, specifically names the California Digital Library among the initiatives it asks the University to support (available at the same url).
[37] Ware, M. and McCabe, M., *The stm report: an overview of scientific and scholarly journal publishing,* September 2009, p. 7. http://www.stm-assoc.org/2009_10_13_MWC_STM_Report.pdf. This figure is based on a 2008 RIN/CEPA study of publishing costs, including print and publisher surpluses but excluding peer review costs. The difference between the RIN/CEPA analysis and actual publisher fees for open access suggests that true article publishing costs may be somewhat lower.
[38] For example, the Biomedical Informatics Research Network no longer provides a repository service for research data. See http://www.birncommunity.org/working-with-birn/faq/
[39] See http://www.cdlib.org/services/uc3/curation/.

[40] See for example, Howard, Jennifer, "Bourbon, Thoroughbreds, and Digital Curation," *The Chronicle of Higher Education*, September 23, 2010. http://chronicle.com/blogPost/Bourbon-Thoroughbreds-and/27107/.
[41] Howard, Jennifer, "Head of British Library Warns of 'a Black Hole' in the Digital Record," The Chronicle of Higher Education, January 29, 2009. http://chronicle.com/blogs/wiredcampus/head-of-british-library-warns-of-a-black-hole-in-the-digital-record/4500
[42] The WAS growth rate does not include CDL's own use of the service for large-scale projects.
[43] Ithaka S+R, *Business Planning Priorities for CDL's Web Archiving Service,* June 2010, p. 3 See footnote: "…if every UC campus purchased a basic service package from Archive-It, the university could build 30 collections with 300 million URLs for around $100,000 a year."
[44] "Bibliographic Services Task Force Report", 2005, http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf
[45] Ibid, p. 4.
[46] National Science Foundation, *Sustainable Digital Data Preservation and Access Network Partners (DataNet).* http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503141.
[47] See https://www.dataone.org/.
[48] See http://www.tib-hannover.de/fileadmin/datacite/index.html.
[49] See founding documents from the 1996-1998 Library Planning and Action Initiative, including THE UNIVERSITY OF CALIFORNIA DIGITAL LIBRARY: A FRAMEWORK FOR PLANNING AND STRATEGIC INITIATIVES at http://www.slp.ucop.edu/initiatives/lpai.html .
[50] Maron, Smith, and Loy. July 2009. *Sustaining Digital Resources: An On-the-Ground View of Projects Today: Ithaka Case Studies in Sustainability*. p. 11.
[51] Based on Internet Archive costs of $.10 per page; a typical book has 330 pages.
[52] For the site for developers and examples of how XTF is being used, see http://xtf.cdlib.org/
[53] For a summary of grants received see http://www.cdlib.org/about/docs/grants_list.pdf.
[54] For the report on mobile services see https://confluence.ucop.edu/download/attachments/26476757/CDL+Mobile+Device+User+Research_final.pdf?version=1