

# UC Merced

## Open Access Policy Deposits

### Title

Next-Generation Technical Services (NGTS): New Modes for Organizing and Providing Access to Special Collections, Archive, and Digital Formats: Final Report

### Permalink

<https://escholarship.org/uc/item/1sz1f058>

### Authors

Dekker, Harrison  
Dooley, Jim  
Kozbial, Ardys  
et al.

### Publication Date

2010-09-10

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27

**Next-Generation Technical Services (NGTS):  
New Modes for Organizing and Providing Access to Special  
Collections, Archive, and Digital Formats**

**Final Report**

September 10, 2010

**New Modes for Access Task Group**

- Harrison Dekker, UC Berkeley*
- Jim Dooley, UC Merced (Steering Team liaison)*
- Ardys Kozbial, UC San Diego*
- Emily Lin, UC Merced (Chair)*
- Eric Milenkiewicz, UC Riverside*
- Lisa Mix, UC San Francisco*
- Sue Perry, UC Santa Cruz*
- Adrian Turner, California Digital Library*

## Contents

<b>Executive Summary</b> .....	<b>1</b>
<b>Establishing the Framework for New Modes of Access</b> .....	<b>3</b>
Reduce Barriers to Access _____	4
Technical Infrastructure _____	5
Coordinated Acquisition Strategy _____	7
Intellectual Property _____	8
<b>Recommendations</b> .....	<b>10</b>
Recommendation 1 / Processing Strategy: “More Product, Less Process” (MPLP) _____	10
Recommendation 2 / Archival Management System: Archivists’ Toolkit (AT) _____	15
Recommendation 3 / Large-scale Digitization _____	19
Recommendation 4 / Digital Asset Management System (DAMS) _____	24
Recommendation 5 / Leverage Micro-services Across the UC Libraries _____	27
Recommendation 6 / Implement a Shared Processing Program _____	31
<b>Appendices</b> .....	<b>35</b>
Appendix 1. Charge	
Appendix 2. NGTS Phase I (Unique Resources Group) Archival Records/Manuscript Collections survey	
Appendix 3. NGTS Phase II (New Modes for Access Group) MARC records survey	
Appendix 4. UC campus digitization technologies survey results	
<b>References and Bibliography</b> .....	<b>35</b>

## Executive Summary

Imagine a UC library with more than half a million volumes, carefully acquired and protected, many of them singular or not commonly held by others in the world. Imagine the contents lining thirteen and a half miles from end to end—more than one Bancroft Library's worth of material—but all of it invisible and inaccessible to users. That library represents the self-reported aggregate of the unprocessed backlog of special collections and archives materials owned by UC.

The UC Libraries have the opportunity to bring such a library online, and make its contents discoverable and accessible to researchers, within this next decade. The UC Libraries also have the opportunity to prevent the creation of such a backlog of newly created scholarly content, much of it in digital format. Collective commitment is needed to invest the time and initial resources to follow through and to capture these opportunities.

These collections must be made discoverable and accessible so that the UC Libraries can begin to measure and understand their usefulness. Otherwise, they only represent the Libraries' best guesses at what UC should be collecting. And only by pushing for the broadest possible access can such measures be well founded and valid.

This report recommends the steps that would bring such a library online and prepare the UC Libraries for their role as stewards and curators of newly created scholarly content. The first section of this report, Establishing the Framework for New Modes of Access, outlines four fundamental areas and the actions that must be taken by appropriate system-wide groups in order to lay the groundwork for new modes of access:

- Reduce barriers to access (Collection Development Committee [CDC]);
- Develop the technical infrastructure for a UC-wide Digital Library (Digital Library Services Task Force 2 [DLSTF2]);
- Coordinate acquisition strategy for primary source materials (CDC);
- Establish system-wide policies about intellectual property rights (Intellectual Property Rights Policy Group).

Our recommendations for eliminating hidden collections and providing timely and effective access for end users are as follows:

### 1. Implement “More Product, Less Process” tactics for processing collections system-wide

**Outcome:** Provide a record for all hidden collections by end of 2012. Eliminate the backlog of unprocessed manuscript collections, archival records, and university archives by end of 2020.

**Cost:**

- Near-term task group: 25% of 5-6 UC Libraries staff members for 2-3 months
- 15% of 1 existing CDL programmer analyst for estimated 6 months
- “Train the trainer” workshop estimated at \$5000 plus travel
- Existing staff for processing

### 2. Support streamlined processing with system-wide use of the Archivists' Toolkit

**Outcome:** Produce collection records and finding aids more efficiently and at less cost. Provide a record for all hidden collections by end of 2012. Eliminate the backlog of unprocessed manuscript collections, archival records, and university archives by end of 2020.

**Cost:**

- 1 FTE project manager to coordinate deployment, training
- “Train the trainer” workshop estimated at \$7,100 plus travel for 50 attendees
- Minimal IT administration
- Existing staff for processing

### 3. Systematically digitize high-use, high-priority collections for access

**Outcome:** Establish an efficient, sustainable system-wide operation and digitize 300-600 l.f. of high priority targets with initial allocation.

**Cost:**

- 1 FTE project manager plus technical specialists from existing staff
- \$200,000 initial allocation plus est. \$21,800/yr preservation storage for 300-600 l.f. digitized content

### 4. Implement a coordinated, system-wide DAMS solution immediately

**Outcome:** Provide all campuses with the means to create, manage, and make accessible digital assets efficiently and at less cost for the near-term.

**Cost:**

- Allocate % time of DLSTF2/implementation team for the near future (2 years)
- Licensing costs and/or development costs to implement available product
- Server storage for digital assets
- Training and maintenance costs

### 5. Leverage UC3 micro-services to manage unique digital assets created or purchased by UC

**Outcome:** Build efficient systems for managing the full lifecycle of digital scholarly content from the beginning

**Cost:**

- Existing CDL/UC3 staffing for development and production
- Infrastructure and overhead costs
- Allocate % time of existing libraries staff for digital curation

### 6. Implement a Shared Processing Program for surveying, processing, and digitizing materials

**Outcome:** Reduce the time to eliminate the backlog significantly (e.g. 50-75%) and make processing, digitization more efficient

**Cost:**

- Reallocation of staff from other library units, e.g. 3 FTE per team
- Allocate % time of existing HOSC and HR staff to reclassify and train staff from other units

## Establishing the Framework for New Modes of Access

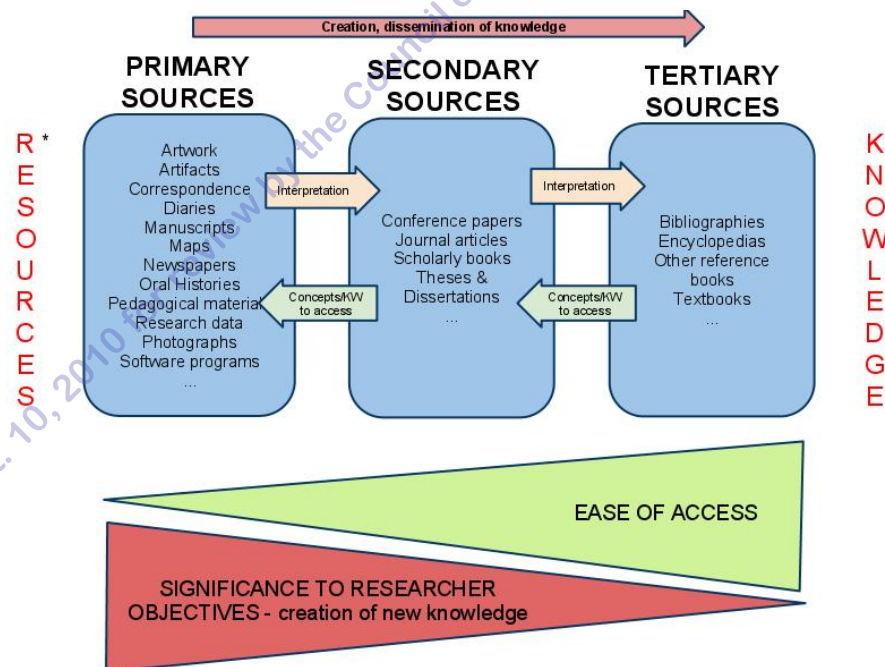
It should be noted from the outset that this Task Group, and even the Steering Team, has worked at getting a handle on the scope of materials and formats to address under our charge. The mental effort of wrapping our heads around what comprises “unique materials” and “special collections, archives, and digital formats” points to the dichotomies that now characterize library collections, and the growing challenge of comprehending and evenly addressing ends of the spectra in our organizations:

- published versus unpublished;
- commonly held versus unique and rare;
- analog versus digital;
- licensed or purchased versus donated or UC-generated.

Returning to the focus of the charge<sup>1</sup>—the needs of the end user—we formulated another construct for thinking about the content, one that considers the overall lifecycle of knowledge creation and the process that scholars and researchers undertake.

Users are concerned with “primary sources” when conducting research. Materials held in special collections—archives, faculty papers, images, rare books and older printed materials—are typically used as primary sources by researchers. While this is obvious in the case of unique materials, older printed materials tend to be used as primary sources too. For example, researchers study a nineteenth-century medical journal as a primary source to learn about issues confronting practitioners at that time, not as a secondary source for current knowledge and practice.

If we use the term “primary source,” the division between analog and digital becomes irrelevant, although more and more users prefer to use digital materials and expect primary source materials to be available digitally.



\* Forms/genres of materials that cut across categories include audio/visual materials, electronic records, images, text, etc.

Figure 1: Model of knowledge creation process based on access to information resources

<sup>1</sup> See Appendix 1: “NGTS New Modes for Access Task Group charge”

Figure 1 illustrates that, for the most part, there is greater ease of access to secondary and tertiary information resources, even though, in the big picture, the availability and accessibility of primary source material is critical to the generation of new knowledge. Overall operations generally have focused on the former of the above-mentioned dichotomies: what is published, and hence commonly held; what is purchased from vendors or publishers; what is analog.

Faced with significantly reduced financial resources and increasing costs for acquiring and managing the exponentially increasing amount of information resources created, the UC Libraries clearly need to establish critical priorities for collecting and for expending resources. Strategic prioritization that will transform operations must focus on the other ends of the spectra: that which is unpublished and thus needs to be surfaced or made accessible, that which is unique, that which is donated or UC-generated, that which is digital.

Until now, these other ends of the spectra have been ancillary to core library operations in terms of decision-making, budgets (often discretionary), and staffing. Thus, the Task Group has identified four key principles that are fundamental to solving the UC Libraries' overarching objectives of increasing efficiency, maximizing resources, and ensuring "broad access over the long term to all of the extraordinary collections of the University of California."<sup>2</sup>

- **Reduce barriers to access to primary source materials**
- **Develop the technical infrastructure for a UC-wide Digital Library**
- **Strategic, system-wide coordination of the acquisition of primary source materials**
- **System-wide policies for intellectual property rights**

These four principles underlie the recommendations presented in this report.

### Reduce Barriers to Access

End users expect immediate access to materials. They have clear expectations for what scholarly resources they can discover and access in the UC Libraries. Essentially, any book or journal article is obtainable. Users do not, however, have a clear picture of what is acquired in other forms and, perhaps even more importantly, why there may be limitations on access.

Physical barriers to access must be considered as well as restrictions due to regulations, confidentiality, donor concerns, etc. In most cases, there are valid reasons for restricting access, though these reasons are not always communicated to the users. While sensitive materials such as confidential correspondence or patient-related materials carry restrictions on access, that does not (or should not) mean that they can never be used.

We propose that CDC take the following actions to reduce barriers to access:

- Expand interlibrary loan of rare and unique materials between UC campuses (that currently includes rare published materials) to include archives and manuscripts collections, adding a Request mechanism to the OAC.<sup>3</sup> New guidelines should be based on the ACRL's Guidelines for the Interlibrary Loan of Rare and Unique Materials.<sup>4</sup>
- Aggressively pursue digitization of primary source materials and make them available online. (See Recommendation 3 for specific CDC actions.)

<sup>2</sup> "Next Generation Technical Services: Changing How We Provide Technical Services for the University of California Libraries," [http://libraries.universityofcalifornia.edu/about/uls/ngts/docs/NGTS\\_scope\\_10april2009.pdf](http://libraries.universityofcalifornia.edu/about/uls/ngts/docs/NGTS_scope_10april2009.pdf), p.1.

<sup>3</sup> Aeon offers a request management software that could be integrated with the OAC. The technical implementation is possible, but policies and procedures will need to be established. See <http://www.atlas-sys.com/products/aeon/>

<sup>4</sup> "Guidelines for the Interlibrary Loan of Rare and Unique Materials," <http://www.ala.org/ala/mgrps/divs/acrl/standards/rareguidelines.cfm>

- Charge a task force to develop system-wide policies for access to rare and unique collections, eliminating unnecessary restrictions. Policies should be informed by the ACRL/SAA Joint Statement on Access to Research Materials in Archives and Special Collections Libraries.<sup>5</sup>
- Ensure that restrictions on access are clearly stated in all delivery mechanisms (catalog record, finding aid, database entry). Access statements should include an end date or requirements for gaining access to restricted material.

If the objective of the UC enterprise is the creation of new knowledge, assert that all content will be made accessible to the public at an identified point in time. Restrictions on access should be the exception, rather than the rule. This assertion needs to be made when accepting or acquiring a collection. However, donor concerns regarding sensitive or confidential material will need to be addressed. Collection managers will need to negotiate workable access agreements in those cases.

### Technical Infrastructure

A commitment to the vision of the University of California Library Collection articulated in the UC Libraries' CDC concept paper<sup>6</sup> calls for a corresponding vision for supporting access to and management of the collection in the 21<sup>st</sup> century and beyond. As the concept paper states, "Emerging technologies promise the potential to create transparent access to and delivery of information across formats and collections and to improve the ability of libraries to coordinate acquisition of information to build the most effective collections."

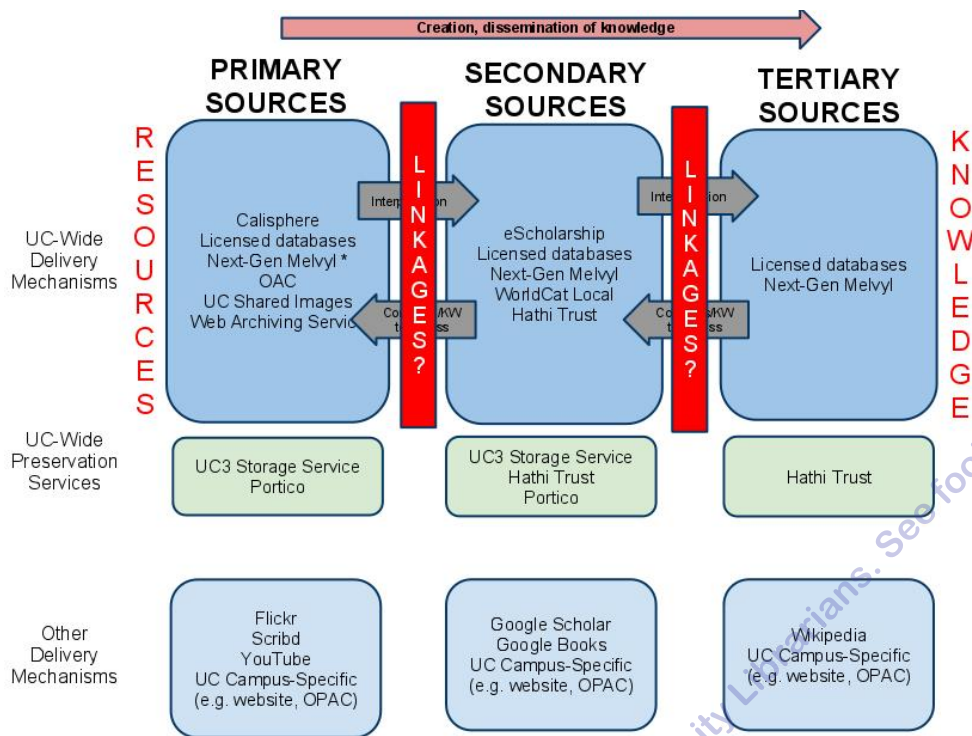
We assert that as discovery and access to information across formats is increasingly digital, there is a critical need to review the UC Libraries current approach to delivering and managing digital information and to develop a comprehensive strategy that fully utilizes state-of-the-art technologies.

---

<sup>5</sup> "ACRL/SAA Joint Statement on Access to Research Materials in Archives and Special Collections Libraries," <http://www.ala.org/ala/mgrps/divs/acrl/standards/jointstatement.cfm>

<sup>6</sup> "The University of California Library Collection: Content for the 21<sup>st</sup> Century and Beyond," [http://libraries.universityofcalifornia.edu/cdc/uc\\_collection\\_concept\\_paper\\_endorsed\\_ULs\\_2009.08.13.pdf](http://libraries.universityofcalifornia.edu/cdc/uc_collection_concept_paper_endorsed_ULs_2009.08.13.pdf)





\* Next-Gen Melvyl does not cover or deliver content at the same level of granularity within this category

Figure 2: Current system-wide services that support the knowledge creation process

Figure 2 displays current system-wide services for delivery and preservation of digital content. Content is often isolated and while some linkages exist, more linkages and services that cut across content categories will need to exist for improved user discovery and access. For example, Next-Generation Melvyl (“Melvyl”) in its current state does *not* cover or deliver content at the same level of granularity across all of the categories shown in the diagram. The delivery mechanisms, or silos, reflected in the diagram are based largely on content production workflows. That is, the delivery of a certain kind of content is based on how it is produced or managed. They do not represent a more holistic view of the desired or optimal discovery experience for the users, to whom content is more important than format<sup>7</sup>, who expect a seamless flow from discovery through delivery, and who prefer immediate online access to content.<sup>8</sup>

Additionally, not all campuses have adopted or been able to leverage these system-wide services; some have relied instead upon local solutions. The Digital Library Services Task Force (DLSTF) report<sup>9</sup> surfaces the unevenness of technological capability and capacity across the ten campuses. It calls for the need to establish “which core set of services and functionality will be maintained centrally” even as the UC Libraries seek to “leverage the strengths of individual campuses both in terms of collections and technical resources and expertise.”

A Library Technologies Advisory Group (LTAG) 2010 snapshot of technology projects submitted to SOPAG reveals that a number of campuses, as well as CDL, are working on development or acquisition of

<sup>7</sup> Schaffner, Jennifer. “The Metadata is the Interface: Better Description for Better Discovery of Archives and Special Collections, Synthesized from User Studies.” (Dublin, Ohio: OCLC Research, 2009). Published online at:

<http://www.oclc.org/programs/publications/reports/2009-06.pdf>

<sup>8</sup> Calhoun, Karen, et al. “Online Catalogs: What Users and Librarians Want.” (Dublin, Ohio: OCLC, 2009). Published online at:

<http://www.oclc.org/reports/onlinecatalogs/fullreport.pdf>

<sup>9</sup> “Final Report - Digital Library Services Task Force,”

[http://libraries.universityofcalifornia.edu/sopag/dlstf/DLSTF\\_Final\\_Report.pdf](http://libraries.universityofcalifornia.edu/sopag/dlstf/DLSTF_Final_Report.pdf)

technologies in support of digital content creation, management, and delivery including, but not limited to digitization equipment, digitization workflow management, digital asset management systems, storage and repository systems, dataset hosting, metadata harvesting, streaming media capability, cloud storage, and user authentication systems. Until now, apart from sharing this information, there has been little emphasis on how these technologies may interrelate beyond the local campus, or the possibility of coordinated planning, evaluation, and development of such technologies.

Combining our collective knowledge and efforts promises a more comprehensive view of system-wide digital library needs and technological solutions, as well as the ability for the UC Libraries as a whole to keep up with the curve. Kruk and McDaniel<sup>10</sup> offer the following as a list of “indispensable services” for the next-generation, semantic digital library:

- Search, browsing, and recommendation services that “exploit interconnected information about resources;”
- Services for augmenting resources, to enable features such as annotation and tagging;
- Dissemination and notification services, to enable access/delivery through mash-ups, RSS and ATOM feeds, and other formats;
- Security and policy assurance services, namely authentication and access control mechanisms;
- Services providing interoperability, such as standards and protocols;
- Preservation services, to ensure versioning and archiving;
- Quality assurance services, to ensure authenticity, quality of metadata;
- Integrated documentation, to help users understand features and how to utilize the resources.

These are services that are the emerging industry standards for Internet content providers, but are not fully identified or addressed in current system-wide digital library services. The UC library system needs a blueprint and a roadmap for delivering these services.<sup>11</sup>

We advocate that the DLSTF2 in concert with LTAG and other appropriate advisory groups should take the following actions:

- Perform a review of system-wide digital library services and identify gaps, overlaps, and opportunities for more effective development of services.
- Create a blueprint and roadmap for system-wide technology infrastructure for digital library content delivery and preservation, both local and central, that also
  - supports programming APIs to allow re-use and re-publication of content, including sharing of objects and/or metadata among campuses as well as with third-party sites such as Flickr, iTunesU, Many Eyes, Scribd, Swivel, and YouTube, and
  - includes mechanisms to capture user-supplied metadata (also known as crowdsourcing) in order to distribute description and promote wide discoverability of materials.
- Address the unevenness of campus technology capabilities; improve communication about and increase collaboration on tools and services under consideration and development.
- Put into place structures and mechanisms that enable rapid development approaches to development of new applications and services, identifying what works and what does not.

### Coordinated Acquisition Strategy

We recommend exploring the potential for increased collective, strategic decision making for acquisition of primary source content. UC acquisition priorities will also inform priorities for processing and digitizing primary source material.

<sup>10</sup> Kruk, Sebastian R. and Bill McDaniel. “Goals of Semantic Digital Libraries,” in *Semantic Digital Libraries* (Berlin: Springer, 2009), p. 71-76.

<sup>11</sup> See Kruk, Westerski, and Kruk, “Architecture of Semantic Libraries” in *Semantic Digital Libraries* (Berlin: Springer, 2009), pp. 77-85 regarding a proposed architecture of semantic digital libraries.

Aligning with the goals of Shared Print, which have thus far primarily focused on secondary source material, and given the realities of space and human resource limitations, the UC Libraries should maximize their resources by having clear priorities and policies about acquisition of primary sources, including what not to duplicate, and what to de-accession. Collection strengths at each campus should be identified and widely communicated.

Clearly articulated, transparent, coordinated acquisition guidelines for primary source collections could result in system-wide efficiencies with the potential to redirect resources where they are most needed. Redundant collecting would be eliminated and campuses would not expend resources competing with each other for unique collections. Further, clear appraisal guidelines (developed with user needs in mind) encourage campuses to concentrate resources on acquiring and processing material that is of most value to users.

System-wide guidelines must, however, recognize that each campus will have its unique priorities, based on factors such as collection strengths and donor relations. Developing donor relations is a key piece of the acquisition process, and a major part of a special collections or archives manager's job. Individual donors often have a connection to a specific campus. If managed and communicated properly, coordinated collection policies have the potential to improve donor relations since many donors are also users. As creators of content, these scholars should understand what would ultimately be of interest to and within the scope of acquisition of the UC Libraries.

We propose CDC take the following actions:

- Revisit bibliographer groups and charges so that there is coordinated acquisition of all primary source material and plans to eliminate unnecessary duplication. CDC should further identify domains for which coordination does not exist.
- Assign to the Heads of Special Collections (HOSC) establishment of system-wide priorities for selection and acquisition of primary source material, articulating and communicating the collecting strengths of each campus. Build upon existing criteria, such as the UC Archivists Council document, "UC Faculty Papers: Identification and Appraisal."<sup>12</sup>
- Standardize or establish best practices at the system level for contracts/deeds of gift for acquisition of primary source material.
- Establish a process to review and communicate these priorities on a regular basis, keeping in mind transitions in staffing, so that selectors/liasons and user communities remain aware of priorities and criteria.

## Intellectual Property

End users expect immediate access to materials. In addition to access, they desire the ability to cite, use, reproduce or display content for research, publication, and instruction. Determining access and rights is often unclear and may vary by campus policy and tolerance for perceived risks.

UC faculty and students are producing new forms of scholarship and research output but require guidance about what kinds of rights can or should be assigned or retained. Libraries are in a position, as trusted stewards of such content, to provide such guidance.

We propose that UC establish or codify system-wide policies about intellectual property rights for primary source material and that an Intellectual Property Rights Policy Group be charged to:

- Standardize system-wide policies about intellectual property rights for primary source material. Ensure that the following types of materials are included: research and pedagogical content generated by UC; grant-funded projects; current holdings. Specifically consider adoption or assignment of Creative Commons or similar licenses.
- Establish risks and system-wide policies for handling orphan works.
- Establish a process to review and communicate these policies on a regular basis.

---

<sup>12</sup> "UC Faculty Papers: Identification and Appraisal," <http://libraries.universityofcalifornia.edu/cdc/ucac/collectingapp2.html>

These actions will enable campuses to reduce resources spent on investigating and tackling legal questions on a case-by-case basis.

Report submitted Sept. 10, 2010 for review by the Council of University Librarians. See footer for priority decisions.

## Recommendations

### Recommendation 1 / Processing Strategy: “More Product, Less Process” (MPLP)

#### Implement efficient “More Product, Less Process” (MPLP) tactics for processing archival and manuscript collections

##### *Rationale*

Archives and manuscript collections, potentially rich in primary source material, make up one of UC's single largest unexposed research resources. UC has made a very large investment in its archives, special collections, and image collections but its “return on investment” cannot be realized until this material can be discovered and used by the academic community.

UC special collections and archives units have over 71,605 linear feet of unprocessed manuscript collections, archival records, and university archives (not including new accessions or pictorial collections).<sup>13</sup> At the current rates of processing and assuming a consistent and focused processing effort, the time it would take to eliminate backlogs on campuses range from 0.2 to 23.6 years (depending on campus).<sup>14</sup> Of the reported 46,662 individual collections maintained across the campuses, 41,806 have collection-level records.<sup>15</sup> Hence, 4,856 lack description and are, for all practical purposes, invisible and inaccessible to end users.

It is simply not sustainable to continue arranging and describing these unique collections across all campuses, using current methods and at the current processing rates. Moreover, reducing massive backlogs will grow exceedingly more difficult while dealing with new accessions.

We propose that the UC Libraries implement, in a coordinated system-wide fashion, the *More Product, Less Process (MPLP)*<sup>16</sup> approach to arranging and describing UC's unique collections, with an immediate goal of providing collection-level descriptions for every archive and manuscript collection in the UC libraries. MPLP comprises a framework of principles and tactics promoting the minimum amount of steps necessary to arrange and describe material sufficiently for use.<sup>17</sup> Some highlights from MPLP that can be employed by the UC Libraries:

- Create collection-level records in the OPAC upon accessioning a collection, rather than waiting until the collection is processed.
- Process at the time of accessioning.
- Reuse any descriptive data that is supplied early in the acquisition process. For example, repurpose information from donors or dealers, reuse accession information.
- Provide data such as box lists online in formats such as PDF, linked to a collection-level record.

Addressing backlogs will also provide a solid foundation to identify and prioritize collections for digitization (see Recommendation 3). Collections must be at least minimally processed in order to pursue digitizing them.

---

<sup>13</sup> Results from NGTS Phase I (Unique Resources Group) Archival Records/Manuscript Collections survey available in Appendix 2. This web-based survey of special collections and archives units was conducted October 5 - 16 by the NGTS Phase 1 UC Unique Collections team. It was distributed to over 50 primary contacts, comprising the UC Heads of Special Collections, UC contributors to the OAC, and to affiliate libraries. Note that not all special collections and archives units reported backlog information -- hence, we believe the total linear feet of unprocessed collections is higher.

<sup>14</sup> In cases where hours to process 1 linear foot were represented as a range, we used the average figure within the range. In cases where hours to process 1 linear foot were not reported, we utilized 8 hours (a commonly-cited average; see Greene-Meissner p. 229).

<sup>15</sup> NGTS Phase II (New Modes for Access Group) MARC records survey responses from UC Heads of Special Collections: see Appendix 3.

<sup>16</sup> Greene, Mark and Dennis Meissner. “More Product, Less Process: Revamping Traditional Archival Processing.” *American Archivist*, 68 (Fall/Winter 2005): 208-263.

<sup>17</sup> It should be noted that some collections will require more intensive processing due to factors such as high demand; donor considerations, the presence of particularly fragile or sensitive materials warranting item-level handling, etc. These should be the exception, rather than the rule.

### Proposed benchmarks for creating collection-level records

Respondents indicated that it takes, on average, 30 minutes to generate a collection-level record.<sup>18</sup> Using 30 minutes as a benchmark and existing staffing levels, 4,856 collection records could be created in approximately 13 months (2,428 working hours), assuming a concerted and sustained cataloging effort.

Campuses utilizing the Archivists' Toolkit (see Recommendation 2) could easily generate an EAD-encoded collection-level record "on the fly," utilizing the same descriptive data that generated the MARC collection-level record.

### Proposed benchmarks for processing collections

Current benchmarks for processing collections range between 2.5 to 8 hours, depending on the campus. We propose adopting a standard MPLP-based benchmark of 4 hours per linear foot across all campuses.<sup>19</sup> (In cases where some campuses are already processing at a faster rate, they should continue processing at that rate). Assuming this new benchmark—and assuming a consistent and focused processing effort, with *existing* staffing levels—the time it would take to tackle backlogs on campuses would decrease.

Based on figures reported in NGTS Phase 1, the amount of time (in years) for campuses to process their backlogs with existing staff compared with the new target benchmark is outlined below.<sup>20</sup>

- **UCB Bancroft Library**
  - Existing benchmark: 27,000 linear feet x 6 hrs / 2,088 = 77.5 years / 17 FTE = 4.6 years
  - New target benchmark: 27,000 linear feet x 4 hrs / 2,088 = 51.7 years / 17 FTE = **3 years**
- **UCD Department of Special Collections**
  - Existing benchmark: 18,650 linear feet x 8 hrs (est., no rate reported) / 2,088 = 71 years / 3 FTE = 23.6 years
  - New target benchmark: 18,650 linear feet x 4 hrs / 2,088 = 35.7 years / 3 FTE = **11.9 years**
- **UCI Special Collections and Archives**
  - Existing benchmark: 2,235 linear feet x 8 hrs / 2,088 = 8.5 years / 1 FTE = 8.5 years
  - New target benchmark: 2,235 linear feet x 4 hrs / 2,088 = 4.3 years / 1 FTE = **4.3 years**
- **UCLA University Archives**
  - Existing benchmark: 1,400 linear feet x 2.5 hrs / 2,088 = 1.7 years / 10 FTE = .2 years
  - New target benchmark: **.2 years (same)**
- **UCM University Archives**
  - Existing benchmark: 20 linear feet x 8 hrs (est., no rate reported) / 2,088 = .1 year / .01 FTE = 10 years
  - New target benchmark: 20 linear feet x 4 hrs / 2,088 = .04 years / .01 FTE = **3.8 years**
- **UCR Special Collections and Archives**
  - Existing benchmark: 3,500 linear feet x 4 hrs / 2,088 = 6.7 years / 2 FTE = 3.4 years
  - New target benchmark: **3.4 years (same)**
- **UCSB Special Collections**

<sup>18</sup> See Appendix 3: "NGTS Phase II (New Modes for Access Group) MARC records survey."

<sup>19</sup> The 4 hours per linear foot benchmark was broadly corroborated by other repositories implementing MPLP-based processing approaches; note that this throughput rate is based on data compiled for processing of large-scale 20th-century manuscript collections and archival records (See Greene-Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," p.253). For other case studies, see Philadelphia Area Consortium of Special Collection Libraries (PACSCL), "PACSCL/CLIR 'Hidden Collections' Project July to December 2009," blog posting online at: <http://clir.pacscl.org/2010/01/13/end-of-year-report-2009/>. See also Linda Sellars, "The True Cost of Collecting: Architectural Collections at NCSU," presentation at CLIR Hidden Collections Symposium, 2010, published online at: <http://www.clir.org/hiddencollections/symposium/NCSU.pdf>

<sup>20</sup> The formula is as follows: start with the existing backlog (in linear feet), multiply this figure by the number of hours needed to process one linear foot. The result is the number of hours necessary to process the backlog. Divide this number of hours by 2,088 (the number of hours in a working year according to the UCOP). The result is the number of working years to process the backlog. Divide this number by the number of FTE to arrive at the number of years it will take the repository to process its backlog.



- Existing benchmark: 1,500 linear feet in backlog x 10 hrs to process 1 linear foot / 2,088 = 7.2 years / 14 FTE = .5 years
- New target benchmark: 1,500 linear feet x 4 hrs / 2,088 = 2.9 years / 14 FTE = **.2 years**
- **UCSC Special Collections and Archives**
  - Existing benchmark: 15,000 linear feet x 3.5 hrs / 2,088 = 25 years / 2.5 FTE = 10 years
  - New target benchmark: **10 years (same)**
- **UCSF Archives and Special Collections**
  - Existing benchmark: 2,200 linear feet x 8 hrs / 2,088 = 8.4 years / .7 FTE = 12 years
  - New target benchmark: 2,200 linear feet x 4 hrs / 2,088 = 4.2 years / .7 FTE = **6 years**

By strategically re-allocating staff from other departments to process backlogs—or adopting an inter-campus processing program—the UC Libraries can further address disparities in staffing levels and backlog sizes between campuses (see Recommendation 6).

### Goals

- Provide collection-level description for every archive and manuscript collection in the UC Libraries, ensuring that every collection is visible to users.
- Increase rates of processing to make primary source materials available sooner.
- Publish collection descriptions and any available discovery tools, such as box lists or inventories in unstructured formats, as early as possible.
- Adopt an iterative approach to the refinement of collection description, reserving more detailed standards when warranted and when time allows for generating better structured data.
- Provide a solid foundation to identify and prioritize collections for digitization.

### What to stop doing:

- UC resources (comprising funds, staff time, supplies,) should be applied across a much wider array of collections, instead of a smaller subset. “We should be paying more attention to achieving basic physical and intellectual control over, and thus being able to permit research access to, *all* our holdings, rather than being content to process a few of them to perfection.”<sup>21</sup>
- Stop waiting to publish collection-level records or finding aids until they are perfect, until detailed container lists are generated, or until the container lists are encoded using EAD.<sup>22</sup>
- Stop wholesale refolding and reboxing of all collection contents when processing.

### Actions

#### Near [2010-1: 3 months]

- Form a short-term working group (5-6 members from archives, special collections, technical services, metadata units, and CDL) to define:
  - System-wide standards for minimal records that will work for MARC and EAD, based on DACS and ISAD-G recommendations. The group should work in concert with the recommendation from the Next-Generation Technical Services Enterprise-wide Task Group for a “good enough” standard.
  - Best practices for inclusion of MARC and EAD records in system-wide delivery systems (such as Melvyl, OAC, and Calisphere) as well as other aggregation services such as OAIster and ArchiveGrid.

<sup>21</sup> See Greene-Meissner, “More Product, Less Process: Revamping Traditional Archival Processing,” p. 237.

<sup>22</sup> Combs, Michelle, et al. “Over, Under, Around, and Through: Getting Around Barriers to EAD Implementation” (Dublin, Ohio: OCLC Research, 2010), p. 12. States: “In the case of providing better access to our collections, the urge to rewrite finding aids (or reprocess collections and then rewrite finding aids) is a huge barrier to providing interim access to the collection descriptions as they are now.” Published online at: <http://www.oclc.org/research/publications/library/2010/2010-04.pdf>.

- System-wide standards for MPLP-based processing levels and recommend practices for implementing particular MPLP-based processing methods.
- A workflow that promotes creating collection-level records in the OPAC upon accessioning a collection, rather than waiting until the collection is processed.

### Mid [2011]

- Implement a project at each campus to create MARC and EAD-encoded collection-level records for each archival record group and manuscript collection in the holdings, using the standards developed by the group described above.
- Provide system-wide training on MPLP principles. (Ongoing)
- Implement MPLP principles for arranging and describing UC's unique collections. (Ongoing)
- Enhance the OAC to support indexing and display of box lists or inventories that are in unstructured data formats.

### Long [2012-Onward]

- Pursue funding to subsidize mass conversion of all remaining legacy print finding aids, box lists, and inventories into PDF format with OCR'd text. The PDF inventories could subsequently be tied to EAD collection-level records.<sup>23</sup>
- Conduct an assessment of MPLP-based approaches to establish revised benchmarks (vis-à-vis processing rates prior to implementing these approaches), and evaluate the effectiveness of the approaches (how many collections are now available). Assess the degree to which additional staff could be redirected to processing efforts. (Ongoing)

## Costs

### Staffing

- For near-term task group assignments, 25% of 5-6 UC Libraries staff members (2-3 months).
- For mid-term and ongoing work:
  - Utilize existing staff and adopt staffing models outlined in Recommendation 6 for ongoing processing work.
  - 15% of 1 CDL Digital Special Collections programmer/analyst staff member time (ca. 6 months) required to upgrade OAC ingest and display systems to support indexing and display of box lists and inventories that are in unstructured data formats.

### Supplies

MPLP recommendations include relying on environmental controls where collections are housed and refolding and reboxing only when original containers are in poor condition. Data about cost savings for archival supplies, based on repositories that have utilized MPLP, do not currently exist but these additional savings are anticipated.

### Training

Training costs to host MPLP workshop are approximately \$5,000 per one day SAA training session. Using a “train the trainer” model, representative participants from each campus could subsequently provide instruction for colleagues.<sup>24</sup>

## Benefits

More primary source material will be accessible to more users sooner. Case studies from other institutions implementing MPLP have indicated that when compared with no access, users prefer timely and immediate

<sup>23</sup> The National Historical Publications and Records Commission (NHPRC) provides basic processing grants that could potentially cover mass-digitization of legacy finding aids. See <http://www.archives.gov/nhprc/>.

<sup>24</sup> Quote provided by SAA to the CDL Digital Special Collections program in July 2009.



access to minimally-processed collections regardless of the descriptive tool or whether a minimal collection-level record or a detailed finding aid exists.<sup>25</sup>

More collection-level records and processing outputs will be gained with the same staffing levels. Reallocating existing library staff and/or utilizing shared processing program approaches will presumably increase outputs.

Addressing backlogs will provide a solid foundation to identify and prioritize collections for digitization (see Recommendation 3). Collections must be at least minimally processed in order to pursue digitizing them.

### **Risks (barriers/obstacles)**

In general, we believe that a major barrier to success is the lack of buy-in on MPLP-based approaches from staff, ranging from repository managers to processing staff. A true shift in culture and mind-set will be required. A strong communication plan, and system-wide training, will be crucial to the success of this effort.

Some UC campus repositories may have concerns with providing access to materials that have potential access restrictions (e.g., student records, medical records, etc.). We suggest that some of these issues can be ameliorated through multiple strategies, such as identifying restricted materials during accessioning, or assessing materials for redaction or restriction when researchers request the files.

Potentially increased demand for reference support as more collections are made discoverable: minimally-processed collections may require more paging and reference requests by researchers, in order to utilize the materials (given that the finding aids may lack detailed descriptions). Additionally, some repositories may have concerns with security risks and potential theft of holdings, if they do not have file- or item-level intellectual and physical control of their collections.

---

<sup>25</sup> McCrea, Donna. "Getting More for Less: Testing a New Processing Model at the University of Montana," *American Archivist*, 68 (Fall/Winter 2005): 208-263.

## Recommendation 2 / Archival Management System: Archivists' Toolkit (AT)

### Support streamlined processing workflows and reuse descriptive data with system-wide use of the Archivists' Toolkit.

#### Rationale

As noted in Recommendation 1, reducing backlogs requires a fundamental system-wide paradigm shift in approaches to processing collections. Archival management systems<sup>26</sup> can significantly facilitate streamlining of processing workflows by promoting reuse of descriptive data, incentivizing cross-sharing of descriptive data between repositories, thereby reducing redundant work and promoting efficiencies.

We propose the UC Libraries develop a coordinated, system-wide implementation as well as a training and support program for UC campuses using the Archivists' Toolkit (AT). The AT already has broad support and uptake within the archival community, including training programs offered by the Society of American Archivists and Society of California Archivists. There is critical mass for use of the software within the UC Libraries: currently, six special collections and archives units have installed and are using the application locally in a production setting, representing 50% of the UC campuses.<sup>27</sup> Two campuses, UC Santa Cruz and UC Merced, are currently assessing the software for local use. Additionally, the CDL is currently hosting an instance of the AT which is available for use by all UC campus repositories.<sup>28</sup>

Using the AT ultimately supports streamlined processing and the efficient and expedient production of descriptive records, thereby promoting access to hidden collections. The AT:

- Produces EAD “on the fly” with the click of a button, eliminating the need for manual EAD encoding—a time consuming process. It similarly exports MARC XML, and METS, MODS, and DC records for digital objects. Base MARC records, in particular, could be generated from the AT by UC special collections and archives staff (archivists, library assistants, student assistants), and those records could in turn be supplemented and edited by technical services staff for loading into Melvyl or local ILS systems. The AT thereby directly supports the goal of creating collection level records for all UC unique collections found in Recommendation 1.
- Reuses data for multiple outputs. For example, accessions information exists in the same database as descriptive information, and data re-use is prevalent throughout the application.
- Aggregates data in one place for managing collections.
- Tracks collection processing priorities and statuses.

#### Goals

- Move from siloed campus implementations of the Archivists' Toolkit toward system-wide, coordinated use of the product, thereby building a shared knowledge base and user community.
- Lower and consolidate costs (IT, staffing, training) across campuses by:
  - Implementing a system-wide instance of the Archivists' Toolkit, with centralized support;

---

<sup>26</sup> Spiro, Lisa. “Archival Management Systems: a Report for the Council on Library and Information Resources” (Washington, DC: CLIR, 2009) CLIR report, p. 1. States: “Archival management systems are a kind of software that typically provide integrated support for the archival workflow, including appraisal, accessioning, description, arrangement, publication of finding aids, collection management, and preservation.”

<sup>27</sup> The following repositories are implementing the software in production: 1) UC Irvine, Special Collections and Archives, 2) UC Riverside, Special Collections and Archives, 3) UC Santa Barbara, California Ethnic and Multicultural Archives, 4) UC San Diego, Mandeville Special Collections Library, 5) UC San Diego, Arts Library, and 6) UC San Francisco, Archives & Special Collections. Given the relatively broad uptake of the AT by UC repositories, and given its range of functionality (in comparison to other options), we are strongly endorsing system-wide use of the AT. We are not aware of a locally-developed campus solution that supports the same range of functions as the AT. There are other somewhat comparable archival management systems available, ranging from commercial to open source solutions (such as Archon or ICA-AtoM). However, the open source solutions, in particular, lack the same range of functionality as the AT. For a further discussion and comparison of open source products, see Spiro “Archival Management Systems: a Report for the Council on Library and Information Resources,” pp. 42-44.

<sup>28</sup> “CDL Hosted Archivists' Toolkit / Archon Service,” <http://www.cdlib.org/services/dsc/tools/at-archon.html>.

- Providing coordinated training on a system-wide level;
- Transitioning all existing UC campus repositories using locally-hosted versions of the AT beginning in the 2012, given that the AT is currently in a significant redesign phase.

#### **What to stop doing:**

- Curtail campus and CDL development efforts to create new, stand-alone archival management systems with limited system-wide application. Divert these resources toward developing a robust solution that is optimized for system-wide use.

#### **Actions**

##### **Near [2011: 3-4 months]**

Under the coordination of a project manager:

- Form a system-wide UC user group for the application.
- Develop and implement a knowledge base (e.g. wiki) for campuses to share information, and provide peer-to-peer user/technical support.
- Develop and implement a training program for UC staff on use of the AT.
- Contract with third-party training organization to host AT workshops. Determine the appropriate model for funding the workshops.
- Investigate implementing a train-the-trainer program, whereby UC campus staff can serve as local trainers.
- Develop help documentation and tutorials to support cross-campus use of the product. Leverage and adapt CDL-hosted AT Service documentation.

##### **Mid [2011: 6-9 months]**

Charge a user group to:

- conduct beta-testing of early iterations of the integrated AT/Archon software.
- conduct an assessment of deployment options for the integrated AT/Archon software, to determine optimal system-wide implementation for the software.
- prepare a recommendation for deploying the integrated AT/Archon software.

##### **Long [2012-Onward]**

- Implement recommendations from the assessment phase.
- Migrate data and transition UC campus repositories to the CDL-hosted AT service.

#### **Costs**

##### **Staffing**

- 1 FTE Project Manager to coordinate deployment, training

##### **Training**

System-wide co-sponsored training session by SAA: \$7,100 for two separate 2-day training sessions, for 50 attendees per session. (\$142 per attendee).<sup>29</sup>

##### **Hardware, Software, and IT Administration**

We anticipate minimal costs. The AT is an open-source product and is available to download for free. Hardware and other material costs comprise hosting of one of the three supported back-end databases (MySQL 5.0, Microsoft SQL Server 2005, and Oracle 10g) and maintenance of that host environment (server

<sup>29</sup> Quote provided by SAA to the CDL Digital Special Collections program in July 2009.

space for the database; support for data backup and recovery; support for network access from the client software to the database).

For the proposed near term goals, significant ongoing hardware, software, or IT staffing costs to implement the software are not anticipated: these costs have already largely been incurred. The costs primarily comprise administration of the database and client software (installation of the database and client software, managing upgrades). For both the campus-hosted and CDL-hosted deployments, it is estimated that IT staffing costs to support the software are fractional, at the level of less than 1% of 1 IT staff member's time (20-40 hours total) for initial setup and configuration of software. Ongoing maintenance of the databases is a marginal cost, but it is assumed that there will be incremental costs based on the addition of more content. IT staff time allocations would rise to accommodate upgrades, migrations, and release packages that particularly pertain to the back-end databases. The AT is not under active development at this time, however, and there have not been any significant updates to the back-end database.

UC campus repositories that are not currently using the AT but are planning to use it in the 2010-2011 timeframe should be directed to use the CDL-hosted AT service to avoid costs involved with implementing the software locally.

For the proposed mid- and long-term goals, the hardware, software, and IT administration costs for implementing the integrated AT/Archon software are unknown at this time. However, it is assumed that these costs would be minimized at the campus level if all campuses utilized the CDL-hosted AT service.

### **Benefits**

The AT was expressly designed to support repositories in gaining greater and more standardized intellectual and physical control over their collections. This has a direct benefit for users: it significantly facilitates the creation of EAD-encoded finding aids and MARC collection-level records. More primary source materials will be exposed to more users sooner.

- Instead of creating MARC records for collections from scratch or based on a reference document such as a draft finding aid, the AT can automatically generate MARC exports with the push of a button.
- In a survey of UC campuses using the AT, two campuses reported that it takes approximately 10 minutes to generate an acceptable MARC record from the application. Name/subject records may be created in a more distributed fashion: authority record creation and maintenance activities could be distributed system-wide.

Coordinated system-wide use of the AT would provide cost savings by combining the work of supporting and maintaining separate local AT instances or maintaining different local systems. It would enable collaboration and leverage expertise in programming, web development, and data management for the system.

Coordinated system-wide training in the form of co-sponsored training sessions through organizations such as the Society of American Archivists (SAA) would help campuses avoid the costs associated with training individual campus staff. Training 50 UC staff members under the arrangement of individual campuses would cost \$16,250 compared with the \$7,100 quoted above.

As part of the Andrew W. Mellon-funded ArchivesSpace project which started in 2010 and is slated for completion in 2012, the software will be merged with Archon, a different open-source archival management system developed by the University of Illinois.<sup>30</sup> The release of the integrated AT/Archon software is planned for 2012. In reviewing the draft technical architecture report and functional requirements for the proposed

---

<sup>30</sup> See ArchivesSpace project information available at <http://www.archivesspace.org/>.

application, we have found that the developers plan to incorporate a number of new features and that it will be more adaptable to a multi-tenant hosting environment.<sup>31</sup>

Implementing a product such as the AT would help campuses avoid costs incurred when developing new, local alternative solutions. At minimum, these include any staff time devoted to programming and developing a local system.

### **Risks (barriers/obstacles)**

Some campuses have made significant investments in building or maintaining other local systems, and developing local workflows built around those systems: for example, UCB's GenDB application and UCLA's SCREAD application. However, we believe there is enough of a critical mass of adoption and cost savings benefits if a majority of UC campus repositories were to use the AT (and forthcoming AT/Archon integrated application).

With a centrally hosted solution, there may be some limits to the extent to which repositories can customize the application for local use. However, we understand that the forthcoming integrated AT/Archon product may be more flexible in supporting multi-tenant usage.

The planned AT/Archon merger will affect implementation. A delay in that project could cause a delay in the proposed timeline.

---

<sup>31</sup> See ArchivesSpace "Technical Architecture Report" and "High Level Functional Requirements" at <http://www.archivesspace.org/>.

## Recommendation 3 / Large-scale Digitization

**Systematically and efficiently digitize high-use, high-priority collections for access to UC primary resources.**

### Rationale

Users expect to find and access digitized content online that can fulfill their research needs. The increased availability of content online will continue to reinforce user expectations that content is accessible—not merely discoverable—online. UC should “enhance public access to UC collections; digitize special collections”<sup>32</sup> in order to transform access to materials that are uniquely held and usually tightly controlled in terms of physical access.

Digitization demonstrably increases the usage of collections.<sup>33</sup> At present, the OAC is the single point of access to the aggregated collection guides of UC primary resource collections. Currently 16,781 UC collections have a finding aid or collection-level MARC record in the OAC. Yet of those collections, only 334, or 2%, have any associated digital items available online. While campuses may have digitized resources that are not available via the OAC, clearly the majority of collections do not offer access to digital copies, in large part because funding for digitization has been discretionary or grant-based. Most of the campuses also have a minimal amount (less than 1%) of their special collections microfilmed, which could offer a surrogate copy of materials to users faced with physical access limitations.

The UC Libraries need to pursue digitization in a more integrated, sustainable manner to build a useful, cohesive UC Collection. Primary resources without online access will remain local, “museum” collections. Thus digitization should no longer be viewed as a marginal activity or enhancement to existing resources and services. Rather, it should become part of the workflow of collection services, along with processing and cataloging. In that light, much as we envision moving other “commodity services” like cataloging to an enterprise level, digitization processes will be more effectively and efficiently managed with coordination and ongoing, strategic investment in developing expertise.

Campuses have made significant investments in developing the capacity for digitization.<sup>34</sup> The UC Libraries, however, can achieve better return on investment if the equipment and skills are more fully utilized by maximizing and concentrating usage, achieving economies of scale, rather than dispersing and duplicating development of capacity on a campus-by-campus and project-by-project basis. A \$5,000 scanner that only scans 1,000 slides in a year is a poor investment considering that commercial services like ScanCafe charge \$0.35/scan; a comparable throughput of 15,000 scans/year on that scanner would represent better return on the investment. In addition, for “many digitization projects, much of the work is roughly 85% repetitive or routine, with 12-15% difficult and <1% very difficult.”<sup>35</sup> Rather than investing in developing expertise at multiple campuses, digitization centers or a service bureau approach would enable lower grade staff to be trained to handle the majority of the work, while concentrating the expertise needed to handle higher-level issues. The staffing models presented in Recommendation 6 could be used.

### Goals

- Establish a UC budget allocation to digitize materials that would provide maximum benefit and access to users. While seed funding may be pursued in order to create a sustainable framework for

<sup>32</sup> “UC Commission on the Future, Working Group Second Round Recommendations.” Published online at <http://ucfuture.universityofcalifornia.edu/>.

<sup>33</sup> For example, the *Avery E. Field collection* at UCR offered only an online finding aid in 2008-2009. Digital images were added in September 2009 and since then the number of views of the finding aid more than doubled. There were 1249 online views of digital images as well as an increase in the number of in-house reference requests.

<sup>34</sup> See Appendix 4: “UC campus digitization technologies survey results.”

<sup>35</sup> Tanner, Simon. “Cost Reduction in Digitisation,” (2006) MINERVA Plus Project. Published online at: <http://www.minervaeurope.org/publications/costreduction.htm>.



large-scale digitization at UC,<sup>36</sup> an ongoing commitment of UC resources will be required for operations.

- Define selection criteria for large-scale, system-wide digitization based on user interests and demand, collection development priorities, preservation needs, and other strategic goals.
- Establish governance, policies, resource/transaction management, and system-wide workflows that will ensure sustainability of digitization of unique materials, ultimately as part of the accessioning workflow. Include provisions for digitization in acquisition/deed of gift language.
- Leverage identified campus capabilities and expertise to maximize investment in equipment and resources as well as throughput and efficiencies.
- Develop benchmarks for digitization costs that can then be included (pre-allocated) in acquisition costs (just as costs for processing may be included in a donor gift).
- Test and refine the use of mass digitization approaches for non-book formats. Evaluate the outcomes, assess the degree of user acceptance of different outputs (based on different levels of scanning quality, etc.), and determine optimal outputs vis-à-vis costs.
- In tandem with an MPLP approach to processing (see Recommendation 1), digitize collections at the aggregate level. Develop the procedures for batch or aggregate-level description of digitized collections.<sup>37</sup>
- Use crowdsourcing in the quality control process and to correct and enhance metadata, specifically to provide item-level metadata.
- Use statistics gathered from online use (views, downloads) and employ user assessment methods to digital collection usage.

#### **What to stop doing:**

- Stop creating the capacity to digitize at every campus when the ability to transport materials between campuses or between campus and vendors makes sourcing just as efficient.
- Stop focusing on providing item-level descriptions and perfect digital objects; describe materials at an aggregate level and focus on making surrogates available online as soon as possible for users to analyze and interpret.
- Stop designing specialized web sites for each digital collection; prioritize exposure via system-wide access points such as the OAC as well as to aggregators and search engines.

#### **Actions**

##### **Near [2010: 3-4 months]**

- Assign CDC to identify initial strategic targets from already-processed collections that are suitable for large-scale digitization: e.g high-usage collections to which UC hold rights or are in the public domain, and have box/series-level processing and/or have been microfilmed. CDC also should establish a proposal process and evaluation criteria for prospective digitization projects.
- Allocate an initial amount of \$200,000 to begin large-scale digitization of identified targets.<sup>38</sup>
- Charge DLSTF2 (comprised of a project manager and team of experts within the system) to oversee large-scale digitization of selected UC primary resources.

##### **Mid [2011]**

The project management team will:

---

<sup>36</sup> Possibilities include NHPRC grants; UNC received \$216,000 from the Mellon Foundation to develop a model for establishing a sustainable, large-scale digitization program.

<sup>37</sup> Weiss, Karen. "EAD (Entire Archives Digitized): Development of a Working Model for Large Scale Digitization," 2009 SAA Research Forum presentation. Slides online at: <http://www2.archivists.org/sites/all/files/KarenWeiss-SAA-ResearchForum2009.pdf>.

<sup>38</sup> The UNC Southern Historical Collections offers a model for establishing a large-scale, sustainable digitization program that was seeded with \$216,000.

- Pursue grant funding to seed and enable the development of a sustainable framework for UC-wide digitization of primary resources.
- Identify digitization service providers and establish workflows, costs, and transaction models based on UC campus digitization technologies survey results (see Appendix 4). Assess how a “centers of excellence” model for digitization can create efficiencies.
- Establish digitization and metadata targets for collections, based on community best practices and existing system-wide specifications (CDL guidelines). Consult with technical experts at identified digitization service providers and the SCP metadata team.
- Establish a workflow and system for managing and tracking scanning work, whereby metadata can be added to digital objects without physical access to the original objects. This can be facilitated by system-wide adoption of a digital object creation utility (see Recommendation 4).
- Determine workflow for how digital assets will be packaged for system-wide delivery mechanisms and other aggregators; proposed targets are OAC and HathiTrust.
- Determine workflow for deposit of digital assets into the UC3 Merritt storage service.
- Define specifications for crowdsourcing capability for the OAC, i.e. the ability to collect comments, annotations from online users to enhance or correct digital objects. Provide these specifications to CDL for implementation.
- Review current and employ appropriate analytics tools for the OAC that will effectively gather data including:
  - Page hits, object views, downloads
  - How long people spend on a page: which parts of a finding aid do users spend time with, do they go back to the front matter from other places in the finding aid?
  - How are people searching descriptive metadata? If they are using keywords, are they getting results?
- Oversee digitization of selected collections using the initial allocation of \$200,000.
- Prepare for and coordinate ingest of digital content into system-wide delivery and preservation services, and to other aggregators.<sup>39</sup>

### Long [2012-Onward]

- As material becomes available to end users, find out from the users themselves if access is easier, if they are getting the information they need to perform their research. Using the usability study expertise at CDL, employ a variety of methods for gathering data about end user behavior, such as interviews, observations, and surveys:
  - The shortest could be a pop up box consisting of 1-2 questions related to an action that the user just performed. For example, if one of the new collection level records is returned after a search, ask the user if that record contains ample information for their research. The user answers yes or no and the pop up goes away.
  - In some of the short pop up boxes (every 20<sup>th</sup>?) give the user the option of taking a longer survey. Ask questions in that survey about what they are doing with information contained in the record or finding aid and what they might need. Citation? Photographer’s name? An easy way to find out copyright or use restrictions?
- Based on results from initial targets and the above-developed processes, continue digitization of UC primary source materials that are outside the scope of current mass book digitization (i.e. Google Books) projects. The project team will develop a detailed financial and staffing model for ongoing operations. The staffing model could directly incorporate elements from Recommendation 6.

<sup>39</sup> HathiTrust staff have been developing specifications for ingest of locally digitized institutional content (Aug 2010). The HathiTrust plans to begin ingest of samples of locally digitized content from Committee on Institutional Cooperation (CIC) institutions in fall 2010 and increase the scale and scope in 2011. See [http://www.hathitrust.org/updates\\_july2010](http://www.hathitrust.org/updates_july2010).



## Costs

### Staffing

- 1 FTE project manager. For mid-term deliverables, assume additional project team members would be allocated from current staffing to incur no new expense.

### Transport

Transportation costs will depend upon type of material and the handling required in addition, of course, to distance. Paper-based collections may be transported to other campuses for digitization via Tricor.

There are some labor costs entailed with inventorying and preparing items for digitization. For the digitization process to be as efficient and streamlined as possible, materials should be grouped by type and items requiring exceptional handling should be identified. Some of these tasks can be incorporated into processing activities described in Recommendation 1.

### Digitization

If we consider labor costs only and do not factor in equipment expenditures, maintenance, training, or other overhead, we can look at comparable costs from other projects.

The University of North Carolina Libraries' large-scale digitization project for the Southern Historical Collection (SHC) hired one half-time graduate student assistant and documented that they were able to complete digitization of 8,434 items (7.5 linear feet) from a correspondence series in roughly 400 total hours. They produced 15,551 color TIFF files at 400 dpi as well as metadata capture at roughly 1.5 min per scan.<sup>40</sup>

The UC Merced Library recently completed a digitization project of the UCSF Speck collection of cholera pamphlets from the eighteenth and nineteenth centuries. Two half-time undergraduate student assistants were employed and under the supervision of a professional were able to complete scanning and uncorrected OCR of 258 items (8 linear feet) in roughly 240 hours. They produced grayscale TIFF and PDF (image over OCR'd text) files of the total of 15,164 pages at roughly 1 minute per page.

The Smithsonian Archives of American Art has employed one full-time digital imaging technician, who scans an average of 1580 images per week, or roughly 53 feet per year. Current production averages 15 seconds per grayscale scan and 18 seconds for art. They will be acquiring a new system that will increase production to approximately 1 second per scan.

If we estimate that the per page labor cost for scanning of paper collections is between \$0.15 and \$0.30, \$10,000 could cover the labor costs for scanning a 16 to 32 linear ft paper-based/manuscript collection.

### Online delivery and long-term storage

The mass-digitization approach proposed will entail minimal metadata creation and processing at the outset. However, it does require that there be a technical way to associate digital objects with aggregate-level records or finding aids, and assumes that there will be a method for enhancing and capturing metadata at a more detailed level after digital objects are made accessible. The project team will need to utilize a digital content creation system (see Recommendation 4) and coordinate with CDL to ensure that the need for such integration, with access mechanisms and preservation mechanisms, is addressed.

Costs for maintenance and storage need to be factored in. As an example, the UNC SHC project documents that the 27.5 linear-foot Thomas E. Watson Papers Collection required 2 terabytes (TB) of storage for the digital files created. The SHC project staff has decided to move to lossless JPEG2000 master files for future projects, reducing overall storage required, but anticipates that it will require 1 TB of archival storage and 1 TB of web (access) storage per year.

### Summary of estimated costs for large-scale digitization of 300 to 600 linear feet of selected UC collections:

<sup>40</sup> "UNC Libraries Project Documentation: Digitization,"

<http://www.lib.unc.edu/mss/archivalmassdigitization/index.html?section=digaccess&page=digitization>

Project management	1 FTE PM plus technical specialists*
Transportation and scanning	\$200,000
Storage	\$21,800/yr (10-20 TB)*

\*For mid-term deliverables, assume imaging and metadata specialists are allocated from current staffing to incur no new expense. Preservation fees are based on UC3 Merritt storage service costs.

### **Benefits**

- Expose primary research material more immediately to a broad audience of users.
- Researchers will immediately view what is in a collection without mediation; users will be directly engaged in describing and enhancing content.
- Increase usage of materials that are tightly controlled.
- Professionals can focus on higher-level reference and research support rather than access and discovery related questions.
- Maximize resources and infrastructure and increase operational efficiencies.

### **Risks (barriers/obstacles)**

Digitization requires an expenditure of resources. On a large scale, this will require ongoing contribution of resources from outside entities or a re-envisioned prioritization in terms of library expenditures.

The success of this recommendation is predicated on more streamlined processing of archival collections (see Recommendation 1). Archivists must be willing to loosen control of both physical and descriptive metadata associated with digital objects in order for mass digitization principles of efficiency to apply.

Curators will also need to build trust in handling of items by others, and in the quality of the output and outcomes. We will need to address the balance between quality and efficiency.

We will also need to manage expectations of users for:

- when a collection is not completely digitized;
- incomplete or incorrect metadata;
- ability to search comprehensively through materials;
- potentially “lower quality” access files.

## Recommendation 4 / Digital Asset Management System (DAMS)

**Implement a coordinated, system-wide solution for creating and managing digital objects.**

### Rationale

Not all of the campuses have a management system for digital objects, but almost every UC campus is digitizing materials and accepting born-digital materials. A way to add metadata, model the content as digital objects, manage the objects, and ultimately provide access to the content is desperately needed. Systems for accomplishing this work are inconsistently available across the UC campuses. In cases where campuses have a system, there is heterogeneity in implementation, ranging from campus-developed utilities to commercial options. The level of investment in these systems varies widely across the libraries.

A common system would generate cost savings in licensing fees in the case of commercial products, achieve economies of scale in terms of administration and training, and enable easy data transfer and shared workflows among campuses. Using CONTENTdm as an example, a group license for all campuses would cost less than the direct licenses already purchased by three UC campuses. A system-wide offering would also help campuses avoid costs by maintaining or developing their own independent systems.

Many campuses have researched both commercial and open source options and have created robust functional requirements lists while they wait for a system-wide initiative. While local or open source systems may be under development, a “turnkey” solution will allow the UC Libraries to proceed in the near term with gaining better intellectual and physical control over existing digitized collections, and importantly, over born-digital backlogs (especially at-risk materials) and new accessions. Moreover, such a solution could be used in coordination with large-scale digitization efforts (see Recommendation 3) so that digital objects ingested in the system from one location could be processed at another location—for example, quality control checked, described, or enhanced—within a workflow.

### Goals

Provide a DAMS solution that can be used system-wide to fit in the digital asset lifecycle between acquisition/digitization and preservation that:

- Handles a variety of digital file formats (image, video, audio, etc.). Can be used immediately to begin to describe and maintain at-risk born digital assets.
- Is interoperable with system-wide preservation and access services, and with third-party aggregators. Digital object exports can be integrated into UC3 curation micro-services, and published through delivery systems such as OAC and Calisphere. Exports can also be integrated into OCLC WorldCat or HathiTrust.
- Can integrate digital objects created with resource records managed in Archivists’ Toolkit.<sup>41</sup>
- Facilitates the creation of digital objects described at an aggregate level (e.g., an entity comprising multiple content files, with a single minimal descriptive record). Handles bulk ingest of files and metadata as well as itemized creation.
- Supports workflow management, so that staff in different roles and from different units may contribute to resource creation or description. Support for shared vocabularies will promote processing efficiency as well as more effective user discovery.
- Supports, exports variety of character encodings for metadata (e.g. multilingual) as well as common data schemas such as Dublin Core and VRA Core. Exports can be converted easily into METS.
- Supports OAI-PMH harvesting.
- Retains the original identity and integrity of collections while allowing them to be searched collectively. This could be done with campus, unit, and collection level branding.

---

<sup>41</sup> For example, the UNC Libraries have automated linking from finding aids to the digital objects being delivered using CONTENTdm; their source code is potentially adaptable to a UC context. See <http://www.lib.unc.edu/mss/archivalmassdigitization/index.html?section=access#shcaccess>

- Supports multiple views to the public, faculty and students, curators and staff.
- Is easy to use/"frustration-free": has reliable and intuitive functionality.

#### **What to stop doing:**

- Stop planning for and implementing multiple, heterogeneous DAMS systems at each campus. Resources could be directed to processing and digitization, making digital content available to end users faster.

#### **Actions**

##### **Near [2010-11: 3 months]**

Charge the DLSTF2 to evaluate available options, including those already in testing or implementation such as CONTENTdm and the UCSD DAMS, to identify a solution that fulfills the above requirements.

##### **Mid [2011]**

Assign an implementation team to oversee implementation at the system-wide level, which should include:

- Licensing/purchase agreement if a commercial system is selected;
- Data migration where necessary;
- Integration with the AT as well as system-wide preservation and discovery layers;
- Account/user management definitions;
- Help documentation and tutorials to support cross-campus use of the product;
- Training;
- Establishing a system-wide UC user group and knowledge base (e.g. wiki) for campuses to share information, and provide peer-to-peer user/technical support; and
- Versioning/maintenance plan.

##### **Long [2012-Onward]**

- Promote and support incremental uptake of the system, across campuses.
- Assess utility of the system in relation to other systems developed.
- Provide recommendations for further development, improvement .

#### **Costs**

Staffing and costs vary widely among the different commercial and open source options. A more detailed analysis of current commercial and open source options is required. Commercial products in particular require annual license fees while open source systems require a larger investment in IT staff to manage and develop.

Citing CONTENTdm as an example of a commercial product, a UC-hosted version (unlimited items, includes 54 client software installations) would cost a \$49,800 license fee, plus \$8,900 annual maintenance fee.

Citing a local system, the UCSD DAMS requires the work of a project manager and 3-4 programmers and has been in development for several years, but its functions and capabilities are more extensive than many off-the-shelf products. It holds the promise for tight integration with other systems such as the Archivist's Toolkit and digital preservation services.

Overall costs, regardless of system implemented, will comprise:

- Project management staff time
- IT staffing for administration and support
- Server storage for digital assets
- Training and user support

## **Benefits**

One system for managing the UC Digital Collection will:

- Provide cost savings by combining the work of supporting and maintaining several different systems.
- Enable collaboration and leverage expertise in programming, web development, and data management for the system.
- Capitalize on single deployment through coordinate system-wide training, avoiding individualized training costs.
- Promote a basic level of uniformity in the structure and encoding of digital content produced by the campuses, thereby promoting efficient exchange of data between campuses and external repositories
- Implement best practices among campuses for efficient workflows.
- Facilitate delivery of digital objects through system-wide delivery mechanisms. More primary source materials will be exposed to more users sooner.

## **Risks (barriers/obstacles)**

The commercial and open source markets for digital asset management systems are changing rapidly. UC3 micro-services, AT/Archon integration, CollectionSpace are examples of tools and services under development that may eventually provide the functions required in this recommendation. There is a need for a near-term solution, however, at this point in time. Deferring the coordinated implementation of a utility that can be used by multiple campuses to wait for other solutions in development will further delay UC's progress in creating and promoting broad access to its digital assets.

Campuses that have already implemented a DAMS or have experimented with solutions might be reluctant to adopt/implement a system-wide solution. Moving to one system requires buy-in from each campus, migration of data from existing systems, and end-of-life plans for those systems. This requires work and investment in the short- to mid-term. We will realize more cost savings as duplicative services are shut down, and as more campuses buy-in and adopt the system. In addition, the system structure may be less flexible than a homegrown DAMS.

Adherence to standards will be essential to ensure a smooth data migration to the next system.

## Recommendation 5 / Leverage Micro-services Across the UC Libraries

**Using the University of California Curation Center (UC3) micro-services as the foundation, develop and implement infrastructure to manage the unique digital assets created or purchased by the UC system.**

### Rationale

As the UC Libraries have extended the boundaries of collecting beyond physical objects to include scholarly digital content, the system as a whole needs reliable infrastructure to manage the latter. In this recommendation, scholarly digital content refers to a wide range of formats from faculty-created raw data to blogs, from web sites to digitized objects that began their lifecycle as physical objects. The micro-services can be employed to manage all data types. Two statements from the UL-endorsed UC Collection Concept Paper are at the core of this rationale. The first statement is this:

In addition, 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. Some of this material will continue to be published through traditional means, but much will be presented in new formats and media. The new scholarly content includes blogs, research data, and a range of materials that will require curatorial engagement throughout their entire life cycle.<sup>42</sup> Providing this curatorial role will ensure long-term availability and help UC address the downstream economic challenges associated with the commercialization of research output.<sup>43</sup>

The second statement is the sixth goal of the concept paper: “Curate and preserve the scholarly output of the University of California.”<sup>44</sup> There is an urgent need in the system for a content management solution for digital objects created in digitization projects and for related born digital objects. We recommend a short-term solution for this digital content in Recommendation 4. As a long term solution, we recommend a combination of UC3’s micro-services for technical infrastructure and UC3’s collaborative working group model that currently includes three and will eventually include all campuses. To date, the three campuses which have been invited and have agreed to participate are UCSD, UCLA, and UCM.

In this new area of managing digital scholarly content, no single institution has expertise, and a new way of thinking about how the UC Libraries work together is required. While new endeavors do cost money, the UC Libraries have the opportunity work collaboratively to create efficiencies from the beginning, rather than waiting until each campus has made an attempt at a solution. Other recommendations in this report are correcting for the result of working separately.

The *Final Report of the Ribbon Task Force on Sustainable Digital Preservation and Access* states:

Effective governance mechanisms are needed to aggregate the collective interest into an effective preservation strategy and, equally important, ensure that the effort and cost of carrying out preservation are appropriately apportioned.<sup>45</sup>

### One Example of Research Output: Data Sets

The importance of data sets as a primary resource in particular disciplines is not new. The broad variety of disciplines now using these data sets as primary resources combined with their availability in electronic formats have changed the way the UC Libraries must acquire, provide access to and manage them. These data sets may consist of information from local, state and federal agencies, or they may be licensed. They may

<sup>42</sup>The content life cycle includes creation, ingestion or acquisition, documentation, organization, migration, protection, access, and disposition.

<sup>43</sup> “The University of California Library Collection: Content for the 21st Century and Beyond,”

[http://libraries.universityofcalifornia.edu/cdc/uc\\_collection\\_concept\\_paper\\_endorsed\\_ULs\\_2009.08.13.pdf](http://libraries.universityofcalifornia.edu/cdc/uc_collection_concept_paper_endorsed_ULs_2009.08.13.pdf), p. 1.

<sup>44</sup> “The University of California Library Collection: Content for the 21st Century and Beyond,”

[http://libraries.universityofcalifornia.edu/cdc/uc\\_collection\\_concept\\_paper\\_endorsed\\_ULs\\_2009.08.13.pdf](http://libraries.universityofcalifornia.edu/cdc/uc_collection_concept_paper_endorsed_ULs_2009.08.13.pdf), p 2.

<sup>45</sup> “Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information,” <http://brtf.sdsc.edu>, p. 2.



contain geographic data, or data created by social scientists. They may be created by UC faculty and researchers, or by interdisciplinary teams made up of researchers from UC and other research universities.

Currently, a large proportion of accessible data sets are not actually held in library collections. Rather, access is provided using licensed access or consortial agreements. Examples include the Inter-university Consortium for Political and Social Research (ICPSR), the Roper Center, and Sociometrics. ICPSR, the largest source of data sets, makes available MARC (and DDI) records for their collection, which have been added to some UC catalogs. In addition, many data sets are published on the web and may be freely accessed by users. Much of this data is published by the government agencies, such as the Census Bureau, the Department of Education, and the Department of Energy, but increasing amounts are coming from non-profits and academia.

Access to data sets in the UC library system shows some of the same problems as other primary source materials: in particular, inconsistent processing, cataloging and acquisition policies. Access problems are exacerbated by some unique characteristics of data. For example, the ability to re-use data requires accompanying technical metadata. Unlike a text in which the words have significant standalone meaning, research data in most cases becomes random noise unless accompanied by basic information about file layout, variable characteristics, collection methodology, and so on. Data are often generated and published in proprietary formats with no guarantee that the software used to create or open the file will be available for future users.

The UC Libraries face an even more fundamental problem if they wish to take on the role of data stewards: lack of staff with adequate training to work with data, and to provide guidance and assistance to data creators and users. If the UC Libraries are to take on the new role of data stewards, they must establish both the human resource and technological capacity to do so and establish the institutional credibility that they are actually up to the task.

### UC3 Working Group Model

To address building efficiencies on the people side of the data management equation, UC3's collaborative model is a good starting point. Since its start-up phase in March 2010, UC3 has used the working group model to make progress in the areas of Technology, Best Practices, Pilots and Marketing. Each campus, plus CDL, brings different experience to the group. Sharing information and comparing notes make the resulting work more complete and credible than it would have been were the campuses or CDL to work alone.

Future plans for the UC3 Working Groups include inviting all 10 campuses to participate and expanding or changing the focus of the groups as needed.

### UC3 Micro-services: Merritt

Development of UC3's micro-services<sup>46</sup> began in January 2010 and these micro-services are the foundation of Merritt. The design of Merritt allows UC3 to respond in a new and efficient way to the ever-increasing number, size, and diversity of content and the ever-increasing diversity of partners, stakeholders, and expectations while simultaneously accommodating the ever-changing technical environment. Because Merritt is composed of small, discrete micro-services it is easier to develop, maintain, and enhance; components are also easier to replace when they have outlived their usefulness.

On September 15, 2010, UC3 is scheduled to deliver approximately one third of the micro-services stack to the UC community. Combined, the following services are the new Merritt Repository.

- An **authorization service** which allows a user to login and use designated services;
- A **storage service** which provides a means to manage digital object files;
- An **ingest service** which provides a means to add new digital content (or a new version) into Merritt for active management;
- An **access service** which provides a means to access a particular object;

---

<sup>46</sup> "Curation Micro-Services," <http://www.cdlib.org/services/uc3/curation/>.

- An **inventory service** which provides a means to associate various types of syntactic, semantic, and pragmatic descriptive information with digital objects; and
- An **identity service**, named EZID, which provides a means to acquire and manage persistent identifiers for resource objects.

These are foundational services that will enable the campuses to efficiently and sustainably manage and preserve their content. The services can be deployed locally or the campuses may use CDL's hosted services. UC3 is anxious to engage with the campuses to better understand how the services can be used and understand the development path that commences after the services above are delivered. When the user deposits an object in Merritt they will be able to retrieve that object via an access system. For example, if a user deposits an object in Merritt, creates a Melvyl or local OPAC record, adds an 856 field to the record, an end user can then retrieve the object. Similarly, if a campus has a special collection that is being served up via the web they can serve up content that is stored in Merritt.

### Goals

- Use the opportunity of a new area of collecting to build efficient systems for managing the full lifecycle of digital scholarly content from the beginning.
- Make system-wide use of the micro-services to maintain digital assets with an initial focus on data.
- Ensure that all staff (technical services, public service, IT) are adequately trained to deal with the new digital scholarly content.

### What to stop doing:

- Eliminate workarounds such as burning a dataset to CD-ROM in order to make more familiar cataloging approaches work.
- Stop building repositories and other data management solutions on each campus without consulting others. Instead, work together as the UC Libraries on the data management issue.

### Actions

#### Near [2010: 3-4 months]

- Use the micro-services proactively at all campuses.
- Develop a cost model for the micro-services.
- Implement a pilot stage for ingesting and providing access to UC-researcher-created data.
- Implement a stage for ingest of digital objects for preservation.
- Bulk-load records into Melvyl for data collections from ICPSR and Roper.

#### Mid-Long [2011-Onward]

- Develop the remaining services in the UC3 micro-services stack.
- Expand the UC3 Working Groups to include all campuses.
- Expand the UC3 Working Groups to include additional areas of work such as data services training for public services staff and policies for inclusion of UC-researcher-created data.

### Costs

#### Micro-services Development and Production

Building on what UC3 has already created in both the working groups and the micro-services offers an economical way for the UC community to participate and have an impact on future development.

In order to run Merritt in full production, UC3 estimates that approximately 1.5 FTE (developer and manager) are needed. In addition to the FTE costs there are additional infrastructure costs (production environment, servers, etc.) and overhead costs (power, etc.).



Development of Merritt during the last ten months is estimated at approximately 4 FTE using a mix of the following staff. Additional costs include infrastructure costs (development environment etc) and overhead.

- Lead designer and architect (PAIV)
- Lead preservation specialist (PAIV)
- 2 Lead developers (PAIV)
- Developer (PAIII)
- User interface developer (PAIII)
- System administration (PAIII)

### Cost Comparisons Between Merritt (UC3) and Outsourced Data Management Providers

Provider	Cost per TB	Services
Merritt	\$1090 <sup>47</sup> storage plus management and development costs already incurred by CDL.	Micro-services. As of Sept 15, 2010: authorization, storage, ingest, access, inventory, identity.
HathiTrust	\$3860 (min) <sup>48</sup> Cost is calculated in GB and decreases by \$.30/GB from 2010-2013. There is a one-time fee of 25% of annual cost for new content.	The infrastructure is made up of five elements: storage, data centers, tape backup, servers and miscellaneous hardware, and staff to oversee and maintain these elements.
Amazon S3	\$1500 <sup>49</sup> for the first 50 TB, costs decrease as TB increase. Cost for Northern California is \$1600 for the first 50 TB.	Storage. Data transfer and other requests are extra.
DuraCloud	Pricing model and services will be available in the 1 <sup>st</sup> quarter of 2011. <sup>50</sup>	

#### Staffing

- Technical services staff time for bulk-loading records into Melvyl for data collections from ICPSR and Roper.
- Allocation of existing staff time to participate in working groups.

#### Benefits

- End user access to digital scholarly content for the long term.
- Building a data management system efficiently and economically from the start.
- Participating in the UC3 Working Groups allows efficiencies through collaboration.

#### Risks (barriers/obstacles)

- Cost is incurred in starting a new collecting area and the services for it.
- Loss of credibility in this domain if initial efforts are not well thought out.

<sup>47</sup> Data are stored at the San Diego Supercomputer Center and the campuses are charged the same. See <http://www.sdsc.edu/services/StorageBackup.html>.

<sup>48</sup> "Hathi Trust," <http://www.hathitrust.org/cost>.

<sup>49</sup> "Amazon Simple Storage Service," <http://aws.amazon.com/s3/>.

<sup>50</sup> "Duracloud," <http://www.duraspace.org/duracloud.php>.

- Participating in the UC3 Working Groups takes resources (people time) away from the campus responsibilities.

## Recommendation 6 / Implement a Shared Processing Program

**Reallocate library staff from units other than archives and special collections for surveying, processing and digitizing materials through implementation of an inter-campus processing program.**

### Rationale

Almost every respondent to the survey of archives and special collections units (done in NGTS Phase 1) expressed the need for more staff to dedicate to processing collections. In order to tackle the collective backlogs, programmatically keep pace with new accessions and at the same time, programmatically move forward with large-scale digitization of collections, the UC Libraries need more flexible mechanisms to amplify existing staff resources. The reality is that UC cannot hire new staff. We propose the following staffing models to strategically supplement existing staffing levels without hiring new staff.

### Goals

Adopt one or a combination of the following staffing models. Each of these models assumes that professional archivists and special collections librarians at each campus would work closely with the processing teams to set priorities, develop processing work plans, and manage donor relations.

**Reallocate staff from departments outside archives and special collections to surveying, processing, and digitizing collections.**

Existing staff from all areas of the library should be repurposed, retrained, and reallocated (including adopting strategies for cross-campus sharing) toward surveying, processing, and digitizing collections. Presumably, some technical services staff time will be freed due to efficiencies realized in processing commonly-held materials and these staff hours could be redirected toward processing and digitizing collections. Additionally, staff from other departments (from preservation to circulation to acquisitions) could be redirected to processing and digitization work. Library Assistants, other paraprofessionals, and students could be trained to perform basic tasks.<sup>51</sup>

**Implement an inter-campus Shared Processing Program**

The program should be somewhat analogous to UC's Shared Cataloging Program or, more broadly, to third-party vendor services such as History Associates. There are two potential models to explore, each with pros and cons. They are not mutually exclusive, and could be used in combination. The team for each model comprises a small group of Library Assistants and other paraprofessionals, trained to survey, process, and digitize materials, plus a professional staff member serving as the supervisor.

- **Model A: Mobile processing teams** between UC campuses in close proximity. Multiple campuses in close proximity would establish and share processing staff. For example, UCSF, UCB, UCD, and potentially UCSC in northern California; and UCLA, UCI, UCR and potentially UCSD in southern California. Campuses participating in this model would designate which staff to share, to populate the core teams. Processing teams could potentially use inter-campus shuttle services, where available.

- Pros

---

<sup>51</sup> UC should consider a more extensive incorporation of student assistants (undergraduates and graduates) into processing programs. The UCLA Center for Primary Research and Training Opportunities provides an excellent model, whereby graduate students are provided with training in archival processing; see <http://www2.library.ucla.edu/specialcollections/researchlibrary/9613.cfm>. Other campuses could conceivably benefit from this effort. Other examples of incorporating student assistants include the University of Chicago's "Uncovering New Chicago Archives Project." See <http://mts.lib.uchicago.edu/project/uncap-proposal.pdf>.

- Does not require shipping/transport of collections. They can be processed on-site at a holding repository.
  - Onsite processing supervisors will be familiar with the collections and available to answer questions on the spot.
  - Mobile processing teams will be able to share expertise with local campus staff.
- Cons
  - This model is only proposed for campuses in close proximity to each other, so not all campuses would benefit from it.
- Other issues
  - Host campus would provide professional staff member (as in-kind) to serve as a supervisor.
- **Model B: Centralized processing teams** at the two UC regional library facilities. Campuses would designate which staff to share, to populate the core team.
  - Pros
    - Service would be available to all campuses.
    - Very strong potential for mass digitization tie-ins (see Recommendation 3). For example, the already-established UC mass digitization framework can be used. In-house scanning capabilities of the regional library facilities can be leveraged.
    - Strong potential for implementing specialized conservation, preservation, and format-specific digitization apparatuses (which would be cost-prohibitive to establish individually at every campus).
  - Cons
    - May requires shipping/transport of collections. Note, however, that many campuses may already have unprocessed collections stored off-site at SRLF and NRLF. Those collections could be prioritized for surveying and processing by shared staff.
    - May place an undue burden on the two campuses that provide professional staff to supervise processing.
  - Other issues
    - Requires two full-time supervisors, one at each regional library facility.
    - Dependent on shared processing protocols (material handling, shipping, processing plans), though one shared staff would assure consistency in procedures across campuses.

## Actions

### Near [2011: 6-9 months]

- Charge a short-term working group to evaluate and recommend the most cost-effective alternative staffing model(s) to supplement existing campus staff levels. If applicable, research staff transportation options, and mechanisms for subsidizing inter-campus travel.
- Charge HOSC to work with local campus HR staff to describe job responsibilities, identify staff, and re-class staff if necessary. Consult with Bancroft Library staff involved with the Bancroft Manuscript Survey Project for job descriptions and staffing models.

### Mid-Long [2011-Onward]

- Train staff on surveying, processing, and digitization.
- Implement new staffing model(s). As an initial approach, begin with surveying unprocessed collections to gain better intellectual and physical control over them. Professional archivists and special collections librarians at each campus can then make informed processing decisions and work plans for the collections, and prioritize them accordingly.

## Costs

We anticipate some costs to implement these staffing models, specifically in these areas:

- Time required by HOSC and local campus HR departments to initiate the process, and time required by reallocated staff for training on surveying, processing, digitization, etc. This would be an up-front cost in terms of staff time.
- Time required by special collections and archives staff to train staff members reallocated from other departments. This would be an intermittent cost in terms of staff time.
- The Shared Processing Program staff model, in particular, may pose some additional cost issues pertaining to subsidizing processing team members' travel between campuses (or to the regional library facilities). This cost would be higher at start-up but costs are anticipated to decrease as processing of backlogs is completed.

## Benefits

The Shared Processing Program in particular offers a flexible solution that allows the UC Libraries to strategically address disparities in staffing levels between campuses. Shared Processing Program teams would reduce the time to eliminate large backlogs significantly. Using data from two repositories noted below, and assuming that a Shared Processing Program team comprises 3 FTE processing at a rate of 4 hours or under, per linear feet (based on Recommendation 1), processing time could be cut by 50-75%.<sup>52</sup>

Based on figures reported in NGTS Phase 1, the amount of time (in years) for two campuses to process their backlogs with existing staff compared with Shared Processing Program staff is outlined below. The formula is as follows: start with the existing backlog (in linear feet), multiply this figure by the number of hours needed to process one linear foot. The result is the number of hours necessary to process the backlog. Divide this number of hours by 2,088 (the number of hours in a working year according to the UCOP). The result is the number of working years to process the backlog. Divide this number by the number of FTE to arrive at the number of years it will take the repository to process its backlog.

- **UCD Department of Special Collections**
  - Current staffing levels: 18,650 linear feet x 8 hrs (est., no rate reported) / 2,088 = 71 years / 3 FTE = 23.6 years
  - With Shared Processing Program staff: 18,650 linear feet x 4 hrs / 2088 = 35.7 years / 6 FTE = **6 years**
- **UCSC Special Collections and Archives**
  - Current staffing levels: 15,000 linear feet x 3.5 hrs / 2,088 = 25 years / 2.5 FTE = 10 years
  - With Shared Processing Program staff: 15,000 linear feet x 3.5 hrs / 2,088 = 25 years / 5.5 FTE = **4.57 years**

The Shared Processing Program model has a very strong potential for large-scale digitization tie-ins (see Recommendation 3).

## Risks (barriers/obstacles)

- Human Resources issues:
  - Challenges with re-allocating or re-classing existing staff and job responsibilities
  - What kind of support is available for inter-campus travel, in particular for Model A? For example, can we subsidize travel costs?
- Requires strong level of trust between campuses.
- Assumes implementation of MPLP principles for arranging and describing UC unique collections.

<sup>52</sup> Appendix 2: "NGTS Phase I (Unique Resources Group) Archival Records/Manuscript Collections survey results."

- Assumes coordinated system-wide use of an application such as the AT to support streamlined workflows and re-use of descriptive data.

Both models may lend themselves better to more homogeneous manuscript collections and archival records.

Report submitted Sept. 10, 2010 for review by the Council of University Librarians. See footer for priority decisions.

## Appendices

### Appendix 1. Charge

<https://confluence.ucop.edu/download/attachments/26214511/NGTS+-+new+modes+for+access+100426.pdf?version=1>

### Appendix 2. NGTS Phase I (Unique Resources Group) Archival Records/Manuscript Collections survey

[https://wiki.ucop.edu/download/attachments/7702742/ngts3\\_spcoll-arch\\_survey\\_report.doc?version=2](https://wiki.ucop.edu/download/attachments/7702742/ngts3_spcoll-arch_survey_report.doc?version=2)

*Raw data (UC special collections and archives units within campus library system only):*

[https://wiki.ucop.edu/download/attachments/26214511/ngts3\\_uclibrary\\_only\\_spcoll-arch\\_survey\\_raw.xlsx?version=1](https://wiki.ucop.edu/download/attachments/26214511/ngts3_uclibrary_only_spcoll-arch_survey_raw.xlsx?version=1)

### Appendix 3. NGTS Phase II (New Modes for Access Group) MARC records survey

[https://confluence.ucop.edu/download/attachments/26214511/ngts-nmag\\_spcoll-arch\\_marc\\_survey.xlsx?version=6](https://confluence.ucop.edu/download/attachments/26214511/ngts-nmag_spcoll-arch_marc_survey.xlsx?version=6)

### Appendix 4. UC campus digitization technologies survey results

[https://confluence.ucop.edu/download/attachments/26214511/UC\\_campus\\_digitization\\_technologies.xlsx?version=2](https://confluence.ucop.edu/download/attachments/26214511/UC_campus_digitization_technologies.xlsx?version=2)

## References and Bibliography

<https://confluence.ucop.edu/display/NGTS/References>

Report submitted Sept. 10, 2010 for review by the Council of University Librarians. See footer for priority decisions.