To: SOPAG

From: Digital Library Services Task Force:
        Diane Bisom, UCR (Chair)
        Kathleen Cameron, UCSF
        Robin Chandler, UCSD
        Robin Dale, UCSC
        Brad Eden, UCSB
        Lynne Grigsby, UCB
        Rosalie Lack, CDL
        Emily Lin, UCM
        Colby Riggs, UCI
        Gordon Theil, UCLA
        John Wang, UCLA

        Sherri Berger, CDL

Re: Interim Report -- Digital Library Services Task Force

Interim Report
Digital Library Services Task Force

Introduction
The Digital Library Services Task Force (DLSTF) has been working since April 2009. This Interim Report presents the Task Force's progress on the charge, initial thoughts, potential areas of recommendations, and fall work plan. The Digital Library Services Task force looks forward to feedback from SOPAG on this Interim Report.

Summary
After hours of discussions, the Task Force has identified three major issues:

1. Collection development issues and policies

    At this time, UC Libraries should move beyond looking at digital initiatives as pilot or experimental. Projects must be evaluated and selected based on established criteria as to how they move the digital library forward, for example by filling a subject gap, complementing other collections (analog or digital), showcasing UC's unique collections, etc.

    It is critical that our concept of the digital collection moves beyond an aggregation of content in a search engine to the reality of making a broad range of digital content available to users. The Task Force believes there is a difference between a "collection of digital initiatives" and a true "digital library." A digital library would furthermore make clear the connections and relationships between digital initiatives—connections that may already exist but have not yet been explored or documented.

    An inventory of current and completed digital initiatives revealed our combined digital assets to be both eclectic and diverse. Some initiatives were undertaken as part of a strategic decision-

making process, but most appear to be opportunistic. In order to move forward rationally, two key questions must be answered: how are these collections/initiatives produced in an administrative sense—how are the decisions made? And what is the process for doing so?

While each UC library has in place collection development policies and long-term strategies for building collections, only some libraries include in them the development of *digital* collections. There is a great need for more strategic direction as we partner with faculty and researchers to collect and create new digital content.

The recently released CDC strategic visioning document "The University of California Collection: Content for the 21st Century and Beyond" (July 2009) articulates a refocused mission for the UC libraries. It envisions a UC-wide collection development strategy and specifically addresses collecting and creating new digital collections in an integrated and collaborative fashion. The fact that these concepts have emerged in the work of both the Task Force and the CDC suggests that discussions need to occur between these two groups and other entities, as appropriate.

In addition to the inventory of the Libraries' digital initiatives, the DLSTF uncovered digital initiatives (and non-digital collections) all over the campuses, in museums, research centers, etc. that compliment the Libraries' collections. This is an area that warrants further exploration.

2. Technical infrastructure

The technical infrastructure for digital initiatives should be based on a life-cycle workflow—defined by the CDC as creation, ingestion or acquisition, documentation, organization, migration, protection, access, and disposition—that allows for the smooth transition of digital initiatives from their place of origin to larger UC spaces for discovery, access, and preservation. The workflow should be in place at both local and system-wide levels.

Every project surfaces new technical and administrative challenges, and thus it is expected that every project will evolve a little differently. That said, however, efforts should be made to establish and put in place a workflow or planning checklist at the beginning of each digital initiative that complements the overall UC-wide effort. All workflows should:

- Use and leverage technologies currently available and allow for forward migration as technologies improve/change.

- Integrate with services such as Next Gen Melvyl, OAC, DPR, HathiTrust that provide user access and long term preservation.

- Consider the Total Ownership Cost (TOC) of available technologies, including hardware, software, labor, maintenance, power, and other associated costs, to create a sustainable technology platform.

- Be documented and shared.

Furthermore, guidelines and best practices need to be established to determine at what point along the path the initiative moves from a local to centralized infrastructure.

At all campuses, the technical infrastructure for creating and managing digital projects needs to be raised up to a certain level (which itself remains to be determined).

LTAG has been asked to review the Task Force's Technology Inventory, and should be asked to further develop and document this area.

The Task Force plans to further discuss and develop these concepts regarding technical infrastructure. We recognize that technology infrastructure is varied across the campuses. We also recognize that working with a centralized or system-wide workflow (if only at points) means local campuses may need to give up custom requirements. Our question now is: can we identify expertise and resources that can be leveraged to reduce the overall cost of and barriers to adding initiatives to the pipeline and making digital assets available?

3. Mainstream Digital Initiative Activities

Each UC campus needs to develop a workflow that:

- Supports a *collaborative approach to selecting* collections that could be digitized.

- Consults with other campuses as materials are being considered for digitization to enable the creation of rich, integrated, and effective digital collections while showcasing and preserving the identity of individual campus collections.

- Creates a high quality digital object, meeting established UC standards for resolution, scanning, and metadata.

- Deposits a preservation copy in an agreed upon repository, such as CDL's DPR.

- Provides a link to the digital representation that is accessible to the user.

- Creates a collection-level record for all digital initiatives with a URL to online access. This record should be included in Melvyl/WorldCat Local. This will provide for faster cataloging and processing.

- More broadly distributes the records and/or content—for example to CDL services such as eScholarship and OAC/Calisphere and third-party vendors such as ArchiveGrid— when time and resources allow and where appropriate.

Discussions with HOTS and the Next Gen Tech Services Task Force need to be instigated, as the processing of digital initiatives should be a major consideration in future technical services discussions, planning, and decisions.

DLSTF: Activities and Progress (March – August 2009)

The Task Force has successfully:

1. Initiated work in each of the four areas of the charge, as noted.

2. Developed a WBS (Work Breakdown Structure) for the Task Force activities (Charge: 1, 2, 3, and 4).

3. Developed a Working Definition for the UC Digital Collection (Charge: 1A):

   The University of California (UC) Digital Collection is a collaborative effort of the UC Libraries to strategically create, manage, preserve and enable reuse of authoritative digital collections. It features selected resources both licensed for and created within the UC system, targeted to the diverse needs of a vast and broad ranging academic community. Using advanced technologies, the UC Digital Collection is integrated with and extends the collections and services of the UC Libraries. (Revised 6/12/2009)

4. Initiated two inventories: *Current and Planned Digital Initiatives by Campus*; and *Inventory of Current Technologies Used to Support Digital Collections-Initiatives by Campus* (Charge: 1B, 1D, 2, 3A, 4B, 4C).

   - Each campus was asked to provide an inventory of current and planned digital initiatives. For each initiative they were also asked to provide the following information: technology used to create the digital content, technology used for providing end-user access, metadata standards used, original format of the content, file format, why/how the initiative was undertaken, and if it was a collaborative project.

   - The granularity with which people defined the term "initiative" varied somewhat, but most defined an initiative as a collection. Examples include local history images, a selection of medieval manuscripts, and newspapers created by Japanese American internees.

   - As of this writing, there were over 300 initiatives, with all campuses reporting. Most campuses have between 2 and 18 initiatives, with UCSD at 57 (many are one-offs) and UCB with well over 200. Some categories and observations follow, with the caveat that the inventories are a work in progress; not every campus has yet reported on every issue, meaning that there may be some variance in the current report. We will build on and complete the inventories in the coming months.

     a. Subject Area

        The subject area of the initiatives is extremely broad, ranging geographically from local to national to global in scope. The majority of the items are humanities/social sciences, with the exception of UCSF collections, as well as a few science-related collections at UCLA, UCM and UCSD.

     b. Technology Creation

        Campuses reported using a variety of scanner and OCR tools. Only three database systems (Webgendb, HP image database, and DAMS) are currently listed.

     c. Technology Access

        The most reported access method was via local campus website (35 instances reported); after that it was through CDL (12), more specifically via OAC/Calisphere or eScholarship. For most in-process projects, technology access was reported as "yet to be determined."

      d.   Metadata

          The most-used standards are Dublin Core and METS. Other standards used include EAD, TEI, VRA CORE, MODS

      e.   Original Format

          Most of the original formats were printed graphic art and/or texts and images (including maps), followed closely by photographs. There were three audio initiatives reported and only one microfilm initiative.

      f.   File Format

          Jpg/tiff was by far the most common format, followed by PDF.

5. Initiated Survey of Digital Library Definitions and Models of other institutions that have a "UC Digital Collection" concept (Charge: 3B, 4).

   One facet of our charge was to explore the notion of and develop a broad definition for the "UC Digital Collection" and its relationship to current/traditional collections. To begin this work, we performed an environmental scan seeking out institutions that had a digital library or a digital collection. By gathering this data, we would acquire a baseline understanding of our peers, enabling our development of the definition of a UC Digital Collection.

   Our survey was international in scope and identified a broad range of definitions and models revealing various configurations of digital content delivery and production services. For example, typically single institutions defined themselves as dedicated to the production, maintenance, delivery, and preservation of a wide range of high-quality networked information resources supporting education. A single institution in the population we surveyed additionally defined themselves as providing leadership in digital library development.

   We also surveyed consortial models, which typically define themselves as collecting, preserving and making available digital content in the support of education, the common good and future generations. A single organization defined itself as providing world-class leadership in innovative use of information and communication technology. Although we did not include the California Digital Library (CDL) in the survey group, CDL would easily be grouped with the consortial organizations, as it supports the assembly and creative use of the world's scholarship for the UC Libraries and provides tools and services enabling the sharing and use of these materials by said libraries.

6. Completed SWOT (Strengths, Weaknesses, Opportunities, Threats (Charge: 1D, 2A, 3B, 3D, 4):

   The Task Force conducted a SWOT analysis in order to help us more deeply understand the strengths, weaknesses, threats and opportunities of creating the UC Digital Collection. The analysis will assist the Task Force in making recommendations regarding collaboration within and beyond UC and in developing a strategy for the "Second Phase."

   First and foremost among the strengths identified was that UC collectively has a rich and diverse collection. Our instruction and research programs rely heavily on our collections. We have a

strong technological foundation in place, including the experience handling and managing large files of catalog records (i.e., Melvyl). In terms of funding, UC has an excellent track record of receiving grants because we are considered a leader among libraries.

Some of the identified strengths, we realized, represented a "double-edged sword." For example, we have a long history of initiating and implementing successful collaborative projects, but we have little experience partnering with UC non-library organizations and private entities. We have a system-wide committee structure in place, but it has not been tested with an initiative such as this one. We have many standards for creating and sharing digital items, but they do not address all material types. Many of our campuses and the CDL have robust technology and infrastructure in place, but the levels of technology varies across campuses. Finally, being as large as we are, somewhere across UC we have expertise in all areas needed, but the trick—especially given the current budget constraints—is to successfully harness that expertise.

Areas of weakness include lack of project management skills; inconsistent user experience/interface skills across campuses; lack of a system-wide authentication system; lack of funding in general and in particular for storage and for digitization; lack of a sustainable model that would make digital projects part of our daily process, rather than one-off projects; and finally, lack of a system-wide mechanism for managing and documenting copyright.

The threats that were identified ranged from the current government lawsuit with Google; lack of legislative support for higher education; need for all ten campuses' buy-in; the current financial situation and the uncertainty of when it will improve; cost savings that may not realized; the complexity of copyright issues; and, finally, the challenges of decommissioning legacy technology and staying ahead given the rate of technology change.

As a system, we recognize several opportunities, many of which address the weaknesses identified above. We have the opportunity to develop of new technology tools; to build capacity beyond the OPAC; collaborate with new partners; improve the end user experience with our systems; and even, as a result, generate new revenue streams. There is the possibility that we may serve new audiences and provide wider access to UC's collections. We may realize significant cost savings from pooling resources and through the development and implementation of more standardized processes, as well as find increased opportunities for funding. In addition, we could build on and improve current system-wide services such as CDL's DPR, OAC, Melvyl/WorldCat Local, eScholarship, and HathiTrust.

This UC digital library initiative should be considered critical in order for us to maintain our respected position in the library world. It will also provide us with an opportunity to show off UC's best and brightest and attract additional faculty and students.

7. Used a number of methods and tools to work as a task force (Charge: 1, 2, 3, and 4).

   - Held 12 conference calls.

   - Used collaborative software: GoToMeeting, ReadyTalk, and Mind42.

   - Established the Task Force Wiki on the UCSF Confluence Wiki site to record agendas, minutes, working documents, background documents, and inventories.

- Used project management strategies and tools, including WBS, SWOT, and inventories.

Fall Work Plan/Next Steps/Continued Discussions

Pending SOPAG's feedback and recommendations on our Interim Report, the DLSTF will continue our discussions, work towards fulfilling our charge, and write the final report. Our plan for the next several months includes the following areas of work, tied to specific areas of our charge:

1. Explore the notion of and develop a broad, overarching definition for the "UC Digital Collection" and its relationship to current/traditional collections.

    - Draft a vision statement exploring "why we are doing what we're doing" and consider what our users/researchers should be able to do within or with these collections.

    - Examine specific collections, such as the Twain papers, to develop case studies for the purpose of understanding production workflow, management, preservation, and delivery.

    - Survey and gather data on all related materials across the spectrum of collections, independent of format.

    - Develop a greater understanding and awareness of the "competition" to our digital library agenda.

2. Identify and make recommendations on the system-wide technical infrastructure that would be needed to create, maintain, manage and support the creation and discovery of the "UC Digital Collection."

    - Establish requirements and develop guidelines for a digital initiative, i.e. provide the requirements for a model project plan and outline what a digital initiative needs to include, such as outcomes and integration into the larger search/discovery space.

    - Recommend standards, best practices, training, documentation, and ongoing funding strategies

    - Identify "in-house" production options.

3. Establish digital collaboration "conversations" and recommend collaboration pilot projects.

    - Identify mechanisms to assess the current state of and progress for goal-setting for UC; where we want to be in 1 year, 3 years, etc.

    - Recommend a process for capturing what is already being produced—UC's own born-digital research assets (such as ETDs).

4. Develop a strategy for creating a "Second Phase" work plan.

    - Share information with the Next Gen Tech Services group as this effort works on related issues, gathers information, and produces inventories.

    - Consider the work of the CDC on the UC Collection Development Policy/Statement as we assess the overall UC collection strengths and weaknesses.

    - Identify additional areas to be investigated by other all-campus groups.