**Final Report**

**Digital Library Services Task Force 2**

May 10, 2011 with November 10, 2011 update to Appendix D

**Table of Contents**

**Appendices**

## Final Report

## Digital Library Services Task Force 2

May 10, 2011

The Digital Library Services Task Force 2 (DLSTF2) is pleased to present SOPAG with this Final Report. The Task Force members are:

Kathleen Cameron (Chair) – UCSF
Todd Grappone – UCLA
Emily Lin – UC Merced
Brian Mathews – UCSB
Eric Milenkiewicz – UCR
John Ober (3/11-4/11) – CDL
Colby Riggs – UCI
Roger Smith – UCSD
Perry Willett (10/10-2/11) – CDL
Sherri Berger – CDL
Felicia Poe – SOPAG liaison

### I. Charge

The goal of the Digital Library Services Task Force 2 (DLSTF2) was to summarize findings and recommend opportunities, processes, and organizational infrastructure for the next steps in advancing the digital library agenda and the "UC Digital Collection." In January 2011 the charge was expanded to include recommendations made by the Next Generation Technical Services (NGTS) New Modes of Access group. The revised charge was as follows:

1. Identify priorities and potential projects for moving forward the agenda outlined in the DLSTF *Final Report*, including creation of the "UC Digital Collection" (defined in the *Final Report* as a subset of the UC Library Collection).

2. Recommend the technical infrastructure needed to support the UC Digital Collection.

3. Recommend options for an organizational infrastructure to support the UC Digital Collection and other collaborative digital initiatives.

4. Consult and communicate regularly with relevant ACGs, NGM, NGTS, and with other UC groups and committees, as needed.

5. Systematically and efficiently digitize high-use, high-priority collections for access to UC primary resources. (NGTS New Modes recommendation 3)

6. Implement a coordinated, systemwide solution for creating and managing digital objects. (NGTS New Modes recommendation 4)

7. Using the UC3 micro-services as the foundation, develop and implement infrastructure to manage the unique digital assets created or purchased by the UC system. (NGTS New Modes recommendation 5)

## II. Strategy

The first Digital Library Services Task Force (DLSTF1) laid significant groundwork in defining the UC Digital Collection and envisioning how it might be supported.[1] In accordance with our charge, DLSTF2 has built on the work of DLSTF1 and NGTS New Modes to recommend implementation paths for the collection. Whenever possible, we have incorporated existing technologies, expertise, and infrastructure at UC, with the aim of advancing the collection with immediate, attainable steps.

Our recommendations are built around an assertion that building, maintaining, and growing the UC Digital Collection is best approached through two related, but distinct, objectives:

1) Provide broad, integrated access to the vast amount of existing digitized content across the University of California libraries.

2) Provide technical and organizational infrastructure to enable the ongoing creation, access (discovery and delivery), management, and preservation of assets in a variety of formats and from a variety of sources managed or brokered by the UC library community.

We present in this report a vision for the overall technical and organizational infrastructure of the UC Digital Collection (starting on page 3) and recommendations for developing that infrastructure through a phased approach. Broadly, we have structured the phases as follows, with more detail and specific steps starting on page 7:

Phase 1: Aggregate and provide access to metadata for the existing digital collections across UC identified and inventoried by DLSTF1.[2] The focus in this phase will be on discovery of existing collections maintained by campuses that have invested locally in an infrastructure to manage and display them.

Phase 2: Expand on the infrastructure developed in Phase 1 to incorporate metadata *and* objects, in order to accommodate those campuses that either do not have a digital asset management system (DAMS) or are ready to migrate from a current local solution. This would also accommodate CDC goals for specific UC-wide digital collections.

Phase 3: Incorporate content from other centrally managed repositories such as OAC/Calisphere, eScholarship, and HathiTrust.

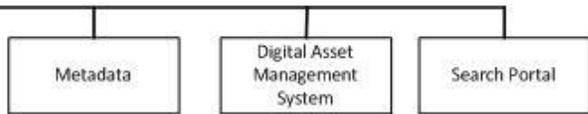Phase 4: Expand the infrastructure to incorporate workflows for born-digital materials.

The diagram below visualizes these phases and suggests in broad strokes the amount of time each phase will require.

---

[1] The UC Digital Collection is defined in the DLSTF1 report as "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."
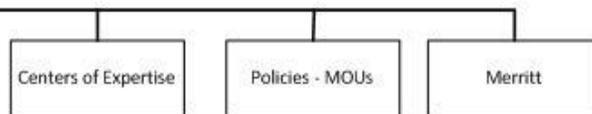
[2] See https://wiki.library.ucsf.edu/display/DLSTF/Current+and+Planned+Digital+Initiatives-Collections+By+Campus.

Months

0            6            12            18            24            30

**Phase 1:**
Harvest metadata from existing UC digital collections

- Metadata
- Digital Asset Management System
- Search Portal

**Phase 2:**
Incorporate new digitized collections

- Centers of Expertise
- Policies - MOUs
- Merritt

**Phase 3:**
Content from other centrally managed repositories

- OAC/Calisphere, eScholarship
- HathiTrust
- Vended/3rd party content e.g. YouTube, Flickr

Digital Library Services Task Force 2
Recommended Phased Approach
Fast Track Timeline
May 2011
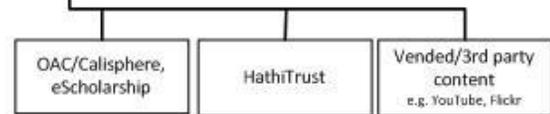
**Phase 4:**
Create a methodology for incorporation of born-digital assets such as datasets.

- Policies
- Common Language

3

Because the structure and details of the later phases will depend on what is initially developed, we have focused our recommendations on the first phase, with an eye towards scalability for later development. Furthermore, we have focused on primary source materials in our recommendations as they comprise the bulk of the existing collections identified by DLSTF1 and provide a good starting point for creating the UC Digital Collection.

### III. Resources and Commitment

DLSTF2 agrees with and reiterates the DLSTF1 assertion that building the UC Digital Collection will require explicit commitments of capital and ongoing resources (staff time, IT infrastructure, and funds) from all UC Libraries.

Additionally, we believe it will require an explicit commitment to widespread adoption, endorsement, and use. This can be achieved through consistent and deep collaboration to both develop and maintain the UC Digital Collection. Collaboration should involve, at minimum, the following tasks:

- Project management for each phase

- Outreach and training for all campuses regarding new processes and procedures

- Development of Centers of Expertise. These will be campuses that have existing equipment and workflows for digitization of specific types of collections (e.g. maps). This is described in more detail below.

- Development of a cost sharing model for UC-wide digital projects and infrastructure support.

- Development of a marketing plan to help drive users to the collections.

- Development of a method for collecting feedback that would inform future improvements to the UC Digital Collection.

### IV. Technical and Organizational Infrastructure

This section presents our long-term vision for the overall architecture of the UC Digital Collection. Our recommendations comprise both *technical infrastructure* (software components and related standards and data flows) and *organizational infrastructure* (commitments, procedures and practices, and repurposed or newly created roles) required to support the collection.

DLSTF2 found it useful to consider the UC Digital Collection from a *collection/object lifecycle* perspective. We have framed our recommendations using the CAMP model, which is defined as: Creation, Access, Management, and Preservation.

### Creation

Our vision for the UC Digital Collection is not prescriptive about where and how objects are created, but instead focuses on the consistency of objects and metadata. Consistency is required so that we can aggregate for discovery and maintain a shared lexicon for rights and preservation.

We recommend the following components related to the creation of digital objects and metadata:

- Selection: Campuses should prioritize collections for digitization based on local needs. CDC may also articulate digital collection priorities at the systemwide level. Both DLSTF1 and NGTS New Modes have made recommendations for the systemwide prioritization of collections for digitization.

- Digitization workflows: The digitization of items and collections will not be prescriptive. Campuses may choose to undertake digitization in-house, using vendors, or with a "Center of Expertise": an identified campus with specialized equipment and processing capacity that can be shared across the system. These Centers could also provide "best practice" consulting on digitization tools, policies and workflows, thus sharing their knowledge across the system and realizing efficiencies in the creation stage. See Appendix D for an initial list of potential Centers.

- Digitization guidelines: The UC Libraries should be working from the same primary set of rules regarding handling by content type. This will create consistency within the UC Digital Collection and build trust with the user community. We recommend the UC-wide adoption of the Guidelines for Digital Images created by the CDL (http://www.cdlib.org/services/dsc/tools/docs/cdl_gdi_v2.pdf). Guidelines for A/V materials are forthcoming from the CDL.

- Metadata: Digitized objects must include adequate metadata to enter the UC Digital Collection. We recommend systemwide adoption of the CDL Guidelines for Digital Objects (http://www.cdlib.org/services/dsc/contribute/docs/GDO.pdf), as this provides baseline discovery and rights metadata requirements, but allows for flexibility if more metadata is desired by the institution.

**Access**

We recommend creating a unified search portal for the UC Digital Collection, in order to facilitate user access to every digital object at the University.

Initially, the system will be a metadata repository allowing users to discover existing primary source material throughout UC through a single access point. Eventually the system will be expanded to include digital objects as well, so that access to objects in the UC Digital Collection will occur in several ways:

- Metadata with links to locally stored collections and items.

- Metadata with links to the OAC, Calisphere, eScholarship, and other centrally managed repositories.

- Metadata and an access copy of the digital object.

Furthermore, DLSTF2 recommends open APIs and flexible workflows so that we might also eventually ingest content hosted on Flickr, iTunesU, and other third-party platforms into the UC Digital Collection.

The task force considered several models of interfaces to distributed collections; a list of these is provided in Appendix C.

**Management**

The NGTS New Modes group recommended implementing a systemwide DAMS. We believe this should form the basis of the UC Digital Collection and how objects are stored and accessed collectively. Campuses with a current DAMS solution will have the option of displaying object metadata in the unified search portal or migrating objects and metadata to the systemwide DAMS. For campuses that do not have a management solution in place, the UC Digital Collection DAMS will provide them with one.

DAMS Product:

In terms of the technology used for a DAMS, DLSTF2 discussed several products. Some we rejected as not meeting a minimum set of consortia requirements (namely, that it is open source, includes open APIs and allows for minimally accepted preservation metadata). The task force selected three DAMS products currently being deployed or developed at the UC Libraries, which we believe warrant further exploration. We have included a brief description of each of these products in Appendix B, in hopes that it will provide a starting point for the next group charged to implement steps toward creating the UC Digital Collection.

Rights Management Framework:

DLSTF2 recommends the creation of a rights management framework requiring systemwide coordination for the UC Digital Collection. These recommendations build upon work completed by the CDL Rights Management Group in 2004 (see http://www.cdlib.org/groups/rmg/).

The framework should address policy and operational issues around rights analysis and clarification, including the formation of requirements for "rights-ready" born-digital content and, ideally, "rights outreach" to authors' activity, in order to encourage maximum access to future materials. The framework must include technical and organizational infrastructure with components that guide rights analysis, record and track rights restrictions, and enable the (automatic) adjustment of access controls at a granular per-object and per rights-change level.

The framework should include, at minimum, the following components:

- *Policies* - including a standard for rights management declarations, expressions, and tracking (note: we believe that existing campus and CDL policies can be gathered, analyzed, and normalized).

- *Operations* - including rights checking; escalation for legal analysis when necessary; and common and consistent rights expressions (in metadata).

- *Rights-based access control advice and methods* - including best practices on the granularity of access-control and tools to use Shibboleth authentication where possible.

To create the framework, we recommend surveying or inventorying existing rights policies and rights operations materials; creating new materials to fill gaps or normalize practices; and establishing reliable legal capacity to assist with rights determinations, when necessary.

Following (or in tandem with) the development of the rights framework, the UC Libraries should conduct a rights and preservation analysis for the  existing digitized collections identified by DLSTF1, so that they are "rights-ready" at the point of metadata harvesting.

**Preservation**

In some ways, preservation is the most complex aspect of the CAMP model. However, UC3 is currently developing a suite of digital curation tools (microservices) with strong preservation components that can readily be incorporated into a new systemwide workflow. This will include depositing all master and derivative digital files into the Merritt preservation repository. The creation of consistent workflows to create preservation-ready objects (format validation, persistent IDs, etc.) will build UC-wide efficiency.

## V. Phased Implementation

This section identifies many of the key factors that will need to be accomplished in a phased approach to developing the UC Digital Collection. We have focused on Phase 1 development; additional work will need to be done to further define Phases 2-4. Efficiencies could be garnered by approaching the development of the proposed phases simultaneously, rather than in a strictly linear fashion, as demonstrated in the proposed timeline on page 3.

DLSTF2 has started a requirements specification, which includes both a) recommended product functions, based those enumerated by NGTS New Modes for a systemwide DAMS solution, and b) a Phase 1 development plan for the UC Digital Collection. See Appendix A for requirements specification.

**Phase 1**

Goal: Facilitate the effective discovery of existing UC resources by harvesting and storing metadata, with active links back to the objects, from the existing, identified UC digital collections.[3]

Actions:

- Verify that enough metadata exists, at either the collection or item level—including preservation metadata—for all objects to be included in the harvest. The CDL guidelines for minimum level metadata required can be used as the basis of this comparison.

- Establish a working group to create the UC Digital Library Rights Management Framework, charged to collect and develop the expertise and resources described above in the Infrastructure: Management section.

- Create a metadata repository and a corresponding access interface. The interface should display rights/access metadata, as well as active links back to the objects wherever they are resolved on local systems. Results should be clustered by type (manuscript, photo, video, etc.) and by campus/location. Both a programmer analyst and a user interface designer will be required to develop the metadata harvest and interface.

- Deploy an enterprise search platform to index targeted collections. We recommend using Apache Solr (http://lucene.apache.org/solr/), which will create and maintain a search across the digital collections, normalizing locally created metadata into a common format. See Appendix A for more information.
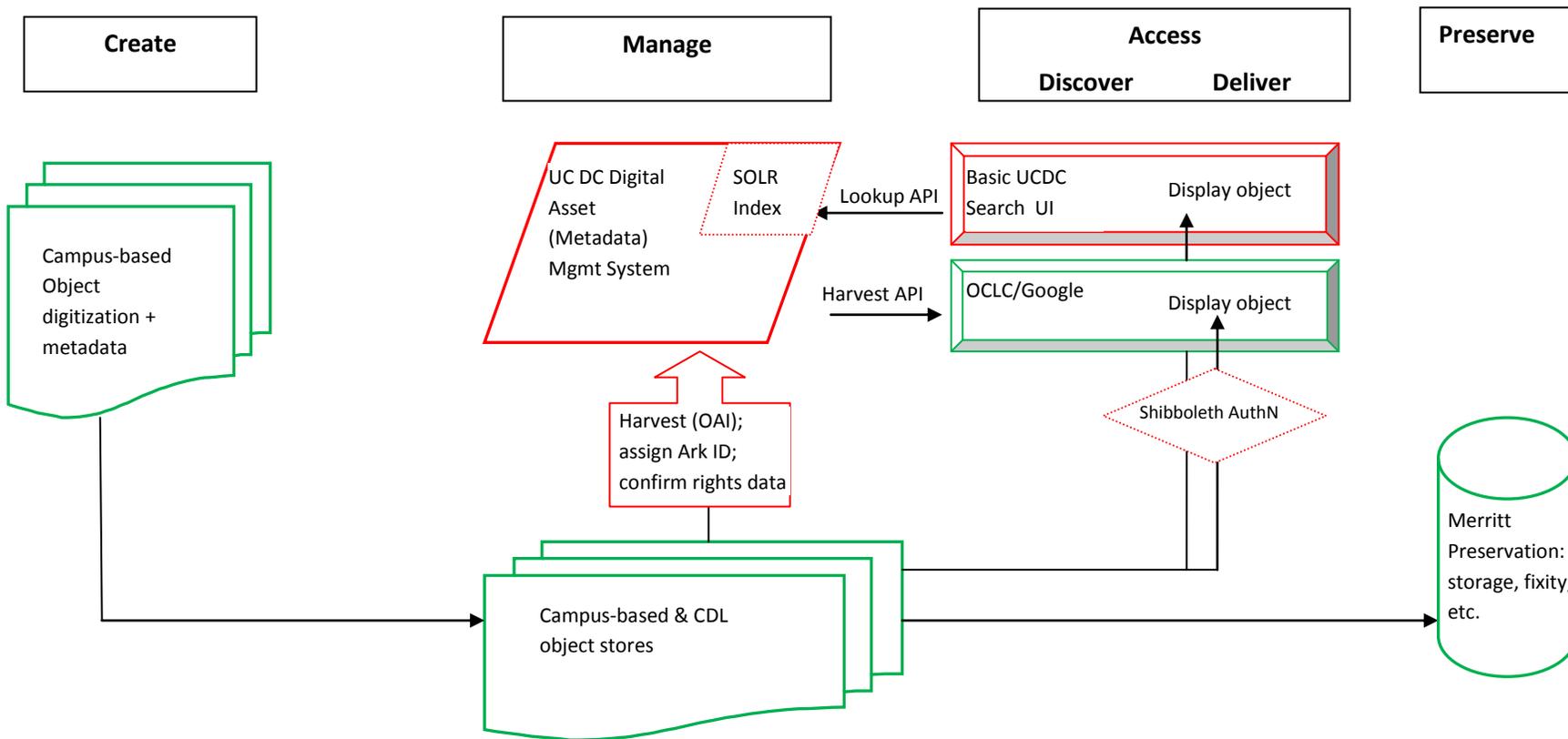
---

[3] For a list of collections identified by DLSTF1, see
https://wiki.library.ucsf.edu/display/DLSTF/Current+and+Planned+Digital+Initiatives-Collections+By+Campus.

- Assign each metadata record a unique identifier, specific to the UC Digital Collection, at harvesting. We recommend using the ARK identifier scheme.

- Create permissions levels so that a campus can manage its—and only its—contributed metadata. We recommend that this authorization functionality be built into the UC Digital Collection DAMS while authentication is provided by Shibboleth (so that content managers can use their local credentials to authenticate).

- Assume that all metadata associated with an object is a public resource, even if the object is rights-restricted. Exposing the metadata for rights-restricted objects poses little risk to the misuse of resources.

- Set read access at the broadest possible level, making the metadata repository open to the public for ease of discovering existing UC primary source material.

- Evaluate and recommend search engine optimization (SEO) strategies as the system is developed.

The diagram on the following page illustrates the proposed architecture of the UC Digital Collection (UCDC) in Phase 1.

# Phase 1 UC Digital Collection (UCDC) Architecture:

## Harvesting metadata from and providing unified discovery and access to existing collections



**Create**

**Manage**

**Access**

Discover　　　Deliver

**Preserve**

Campus-based Object digitization + metadata

UC DC Digital Asset (Metadata) Mgmt System

SOLR Index

Lookup API

Basic UCDC Search UI

Display object

Harvest API

OCLC/Google

Display object

Harvest (OAI); assign Ark ID; confirm rights data

Shibboleth AuthN

Campus-based & CDL object stores

Merritt Preservation: storage, fixity, etc.

Note: Use existing creation tools (e.g. Arch.Toolkit; WebGendb); phase 2 could outsource to "Center of Excellence"

Note: UCDC DAMS populated from harvestable extant systems; UCDC DAMS toolkit may be adapted from existing (UCSD DAMS; CDL's OAC; Fedora; Islandora); requires purpose-built harvesting process/tools that includes Ark-ID assignment and rights metadata confirmation.

Note: Assumes discovery from external sources (Worldcat/Google) AND a newly built/adapted discovery/delivery interface (e.g. from CDL's OAC)

Note: Preservation copy is separately maintained in phase 1; could serve as access copy in phase 2

Green = existing component

Red = new (or adapted) component

9

**Phase 2**

Goal: Expand upon the Phase 1 development effort to accommodate the inclusion of new digitized collections—thereby meeting local campus needs and CDC priorities for UC-wide collections. Campuses that lack a digital asset management system (DAMS) or are ready to migrate from existing solutions (such as ContentDM) could begin participating in the UC Digital Collection.

Actions:

- Create and implement a workflow for metadata and objects: ingest, storage and access. This will require the development of a robust management system to accommodate the entire digital life cycle (creation, access, management, and preservation).

- Every object will be assigned an ARK at ingest, deposited into the repository, and preserved in Merritt. There needs to be accommodation for grouping objects by collection.

- Develop a cost sharing model for UC-wide digital projects.

- Establish Centers of Expertise for digitization within the UC Libraries (see Appendix D). We recommend a system whereby materials are shipped via TriCor to the campus with the required equipment and expertise. This will take advantage of existing investments and resources.

- Develop a simple web-based tool for creating descriptions and rights metadata to be associated with digital objects.

- Include a method for incorporating descriptive metadata through crowd sourcing. DLSTF2 uncovered the following projects currently being undertaken at UC; these should be investigated to determine the best methodology for crowd sourcing:

    o UCLA is working on a project using crowd-sourced metadata.
    o UCSC's Grateful Dead Archives will be using Omeka to allow for public input.
    o UC3 is building an annotation microservice that may provide a modular solution.

- Create a "gatekeeping" role for monitoring crowd-sourced metadata.

- Include version tracking as part of the expanded workflow: masters and derivatives, associated metadata.

- Enable the display of metadata and thumbnail images for objects stored in the system.

- Enhance the interface to accommodate local branding needs.

**Phase 3**

Goal: Incorporate content from other centrally managed repositories such as OAC/Calisphere, eScholarship, and HathiTrust.

Actions:

- Create APIs to accept UC content that is stored in these repositories, as well as potentially other third-party systems like Flickr. Revisit the harvesting mechanism for possible use in collecting metadata from additional sources.
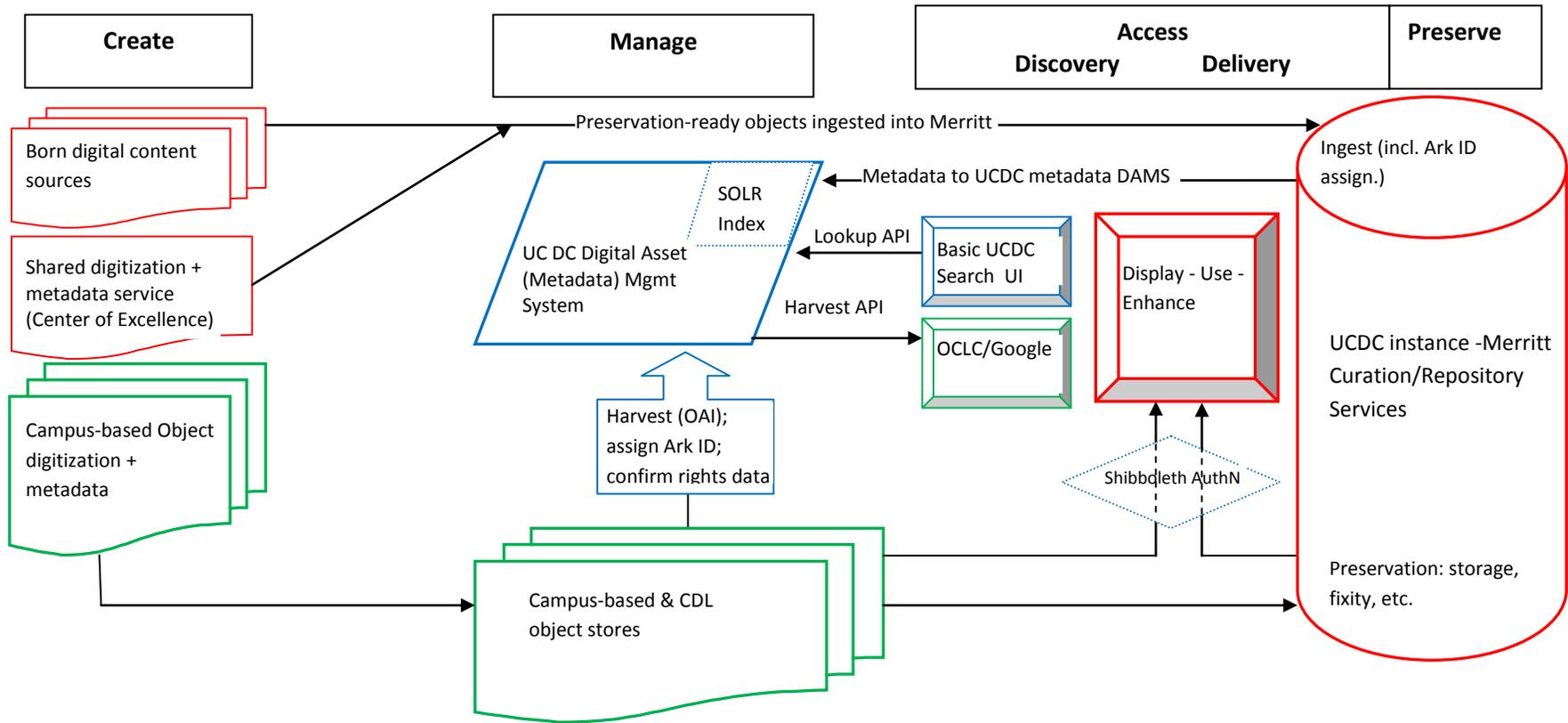
**Phase 4**

Goal: Create a methodology for incorporating born-digital assets such as datasets. The UC Digital Collection architecture is non-prescriptive and will be able to accommodate multiple file types and metadata. The UC3 group has made significant progress in this area and the pending Data Management Plan tool would be integral to this workflow.

Actions:

- An initial focus on defining a service model based on the Data Life Cycle model, as created by the DDI Alliance (http://www.ddialliance.org/) could help determine any additional tools needed for the management of research data.

- Collaborate with campus records management groups to establish a coordinated upstream solution.

- Establish criteria for both selecting and providing levels of access to born-digital assets in the UC Digital Collection.

- Incorporate the CDL's Data Management Plan initiative.

- Create common language, metadata requirements, and management policies.

See below for a chart showing the proposed architecture of the UC Digital Collection as it evolves in phases 2-4.

# Phase 2 UC Digital Collection (UCDC) Architecture:

## Shared/centralized components, expanded content types and sources



| Create | Manage | Access | | Preserve |
|---|---|---|---|---|
| | | Discovery | Delivery | |

**Create components:**
- Born digital content sources
- Shared digitization + metadata service (Center of Excellence)
- Campus-based Object digitization + metadata

**Manage:**
- UC DC Digital Asset (Metadata) Mgmt System
- SOLR Index
- Harvest (OAI); assign Ark ID; confirm rights data
- Campus-based & CDL object stores

Preservation-ready objects ingested into Merritt

Metadata to UCDC metadata DAMS

Lookup API — Basic UCDC Search UI

Harvest API — OCLC/Google

Display - Use - Enhance

Shibboleth AuthN

Ingest (incl. Ark ID assign.)

UCDC instance -Merritt Curation/Repository Services

Preservation: storage, fixity, etc.

Note: Shared digitization capacity "Center of Expertise" + born digital sources.

Note: UCDC DAMS extended with metadata from shared digitization and born digital (metadata harvested from Merritt where objects are deposited).

Note: Assumes discovery from external sources (Worldcat/Google) AND a UCDC discovery/delivery interface; Delivery from either campus access system or UCDC Delivery (Merritt/XTF) service. Assumes extensibility for additional user services (e.g. annotation)

Note: Preservation copy for campus-based objects; single-copy (preservation & access) for Merritt objects.

Green = existing component
Blue = phase 1 completed component
Red = new (or adapted) component

**Appendix A:**

**Requirements Specification for the UC Digital Collection Infrastructure**

a) <u>Product Functions</u>

*Minimal requirements identified by DLSTF2:*

- Assign a persistent identifier (PID) to each asset upon ingest.
- Store digital objects, including related metadata.
- Allow for discovery and display of digital objects.
- Have flexible user roles and permissions.

*Requirements identified by NGTS-New Modes:*

- Handles, supports, and exports a variety of digital file formats (*e.g., image, video, audio*). Examples include:

    - TIFF (graphic materials, textual documents)
    - ALTO (full text transcriptions of textual documents)
    - TEI (full text transcriptions of textual documents)
    - PDF/A (full text transcriptions of textual documents)
    - HTML/XHTML (full text transcriptions of textual documents)
    - BWF/BWAV (audio)
    - AIFF (audio)
    - MXF (video)
    - AVI (video)
    - MOV (video)

- Handles, supports, and exports a variety of character encodings for metadata as well as common data schemas. Examples include:

    - Dublin Core
    - VRA Core
    - PBCore
    - Premis
    - METSRights
    - METS
    - MARC
    - EAD

- Interoperable with systemwide preservation (UC3 Merritt) and access services (OAC, Calisphere).

- Interoperable with third-party aggregators (OCLC WorldCat, Hathi Trust).

- Integrates digital objects created with resource records managed in tools such as the Archivists' Toolkit, Archon, and the to-be-released ArchivesSpace.

- 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 associated metadata.

- Handles itemized creation.

- Supports workflow management that allows staff in different roles and from different units to contribute to resource creation or description.

- Supports shared vocabularies to promote processing efficiency as well as more effective user discovery.

- Supports OAI-PMH harvesting.

- Customizable display that allows the original identity and integrity of collections to be maintained (*e.g., campus specific branding*).

- Supports multiple views for the public, faculty/students, and curators/staff.

- Easy to use, reliable, and intuitive interface that requires limited start-up and ongoing training.

b) Phase 1 Development Plan

As described above, DLSTF2 recommends a staggered implementation approach to the UC Digital Collection. To get up and running quickly, the initial focus will be on aggregating and providing access to metadata for existing digital collections across UC. This will improve discovery by our users, as well as accommodate institutions that have invested in a local infrastructure to support digital collections. It will also set the state for later expansion, namely in order to:

- Support digital objects and provide a DAMS solution for campuses that do not have one or are ready to migrate from a current local system (phase 2).

- Incorporate additional sources and formats such as centrally managed repositories (phase 3) and born-digital content (phase 4).

In order to facilitate the goal of a unified discovery interface for UC Library digital collections, DLSTF2 recommends the deployment of an enterprise search platform. Specifically, DLSFT2 recommends deployment of an Apache Solr (http://lucene.apache.org/solr/) project to index and provide a unified search for targeted collections from each existing UC digital collection. Europeana (http://www.europeana.eu/) is an example of a project that is using Apache Solr to provide access to distributed collections.

We propose a 12 month implementation process that will focus in the first half on rapid development, and in the second half on the user interface and metadata refinement.

This solution will have two component parts: a metadata harvester and a unified index. The harvester will continuously pull metadata from participating Digital Asset Management (DAM) systems, normalize the metadata to MODS, and store the combined result in a unified index. This unified index will provide a user interface to enrich and contextualize a faceted search and discovery experience. Further, we propose a series of search engine optimization (SEO) strategies to enhance all of the metadata and contextual information into a commercially friendly search context to further refine the search and discovery experience. This will ensure that search engines such as Google will be better able to find, crawl and

index our pages resulting in a higher level of visibility and increase traffic. Given that large numbers of traffic to UC Libraries web sites come from organic web searches, we must ensure that all users can find the rich primary sources that already exist. Within the interface, any search results will point back to corresponding records in the source DAM systems, until a time when those digital objects may be migrated to the new systemwide DAMS.

*Anticipated Resources Needed:*

- Funding for an (estimated) 6-month Programmer Analyst to develop a working application as well as a server to host the service.

- Funding for a User Interface Developer. We also recommend a complete search engine optimization (SEO) evaluation and recommendation for the service as it is developed.

- Each campus with a DAM will need to dedicate an onsite technical liaison to the project as local collections are indexed.

**Appendix B:**

**DAMS Products Under Consideration**

DLSTF2 discussed several products for a systemwide DAMS. We selected three DAMS products currently being deployed or developed at the UC Libraries, which we believe warrant further exploration.

1.  UCSD DAMS

The UCSD DAMS product is just part of a very deep workflow and governance commitment that has been made over the last 10 years. The metadata storage methodology is to translate any incoming metadata to RDF and store in a triple store.  This provides an enormous amount of flexibility to consume and share any metadata standard as needs arise.  Scientific data sets could be added with other ontologies and make them available for users.  The general user interface is slick.

2.  Fedora:

(From the Fedora website; http://www.fedora-commons.org/): Fedora (Flexible Extensible Digital Object Repository Architecture) was originally developed by researchers at Cornell University as an architecture for storing, managing, and accessing digital content in the form of digital objects. The software defines a set of abstractions for expressing digital objects, asserting relationships among digital objects, and linking "behaviors" (i.e., services) to digital objects and provides a core repository service (exposed as web-based services with well-defined APIs). In addition, Fedora provides an array of supporting services and applications including search, OAI-PMH, messaging, administrative clients, and more. Fedora provides RDF support and the repository software is integrated with semantic triple store technology, including the Mulgara RDF database. Fedora helps ensure that digital content is durable by providing features that support digital preservation.

UCLA Library has undertaken a project to migrate from their current DAMS to join an open source digital library project called Islandora.  Islandora uniquely combines the Drupal (http://drupal.org/) and Fedora open software applications to create a robust digital asset management system that can be fitted to meet the short and long term collaborative requirements of digital data stewardship. This project will provide a repository and metadata development platform for several campuses to share.

3.  Merritt

Merritt (http://merritt.cdlib.org/) is a new cost-effective repository service from the University of California Curation Center (UC3) that lets the UC community manage, archive, and share its valuable digital content. Merritt features:

- an easy-to-use contributor interface for deposit and update
- access via persistent URLs
- tools for long-term management
- permanent storage

The UC3 group is currently testing a beta prototype of an XTF front-end that could be used as a general user interface. (OAC, Calisphere, and eScholarship are all examples of projects that use XTF.) It may be possible to deposit metadata and objects into Merritt and publish to one of the existing CDL services such as eScholarship.

**Appendix C:**

**Models for Distributed Digital Collections**

Within UC:

- Online Archive of California (http://www.oac.cdlib.org/)
- Calisphere (http://www.calisphere.universityofcalifornia.edu/)

Outside UC:

- Digital Public Library of American planning initiative (http://cyber.law.harvard.edu/dpla/)
- Europeana (http://www.europeana.eu/)
- HathiTrust (http://www.hathitrust.org/)
- University of Wisconsin Forward (http://forward.library.wisconsin.edu/)
- Matapihi (New Zealand; http://matapihi.org.nz/)
- American History Online (http://www.americanhistoryonline.org/)
- Bibliotheca Alexandrina (http://www.bibalex.org/)
- Northwest Digital Archives (Orbis Cascade Alliance, http://nwda.orbiscascade.org/index.shtml)
- Kentuckiana Digital Library (http://kdl.kyvl.org/)
- Smithsonian Institution Collections Search Center (http://collections.si.edu/search/)

## Appendix D:

## Centers of Expertise

DLSTF2 recommends establishing "Centers of Expertise": identified campuses with specialized equipment and processing capacity that can be shared across the system. These Centers could also provide "best practice" consulting on digitization tools, policies and workflows, thus sharing their knowledge across the system and realizing efficiencies in the creation stage.

DLSTF2 identified the following Centers of Expertise. Please note that a more comprehensive list was created as part of the NGTS effort.

| Formats | Expertise | Digitization Specs | Conversion Services |
|---------|-----------|--------------------|---------------------|
| Books (page images +OCR) | CDL, UCM, UCSD | | UCM |
| Books (encoded text) | UCD, UCSB, CDL,UCSD | | |
| Manuscripts | UCM, UCSD | | |
| Finding Aids | | | UCM |
| Mfilm, Mfiche | | | UCM |
| Newspapers | UCSF, UCR | | UCR (CBSR) |
| Photographs (special formats: glass plates, slides, stereographs) | UCSD | UCR | |
| Posters | UCSD | | |
| Maps | UCLA | | |
| Sheet Music | UCLA | | |
| Art works | UCM | UCR | UCM – possibly |
| Audio (special formats: 45s, 78s, cylinder recordings, audio tapes) | UCSB,UCSD | | |
| Video | UCSD | | |
| Scientific Data | | | |