Application for the 2016 Larry L. Sautter Award for Innovation in Information Technology

Submitted by: Sherri Berger, Product Manager, California Digital Library (UCOP)

sherri.berger@ucop.edu | 510-987-9699

<div align="center">

The UC Libraries Digital Collection Project:
Revealing the University's Hidden Treasures

</div>

The Problem

The ten libraries of the University of California (UC) system hold a dazzling array of unique materials, including photographs, newspapers, maps, oral histories, correspondence, government publications, historical records, works of art, films, ephemera, and more. Documents of world events and expressions of human culture, these materials are essential items of study for scholars, students, and the general public—and in many cases, the UC campus libraries hold the only existing copies in the world.

Unfortunately, until recently, many of these vast and valuable resources were effectively invisible to researchers because they were not available online. Why? For years, all of the UC Libraries had been digitizing their remarkable research materials. But several of them lacked the technology needed to support—robustly, routinely, and at scale—the next steps: storage, management, description, and public access. This lack of a post-digitization "pipeline" had slowed or even stalled the digital collections building program on almost every campus.

Project Goals and Timeframe

The UC Libraries Digital Collection (UCLDC) Project was a two-year initiative to establish the technical pipeline needed for all UC Libraries to create, manage, and make accessible their digital assets efficiently and at less cost. In July 2013, following four years of systemwide planning, a project team based at the California Digital Library (CDL) at the Office of the President and IST Research Technologies at UC Berkeley, and in close collaboration with multiple staff members at all ten campus libraries, embarked on the project. (See page 5 for the full roster of participants.) In September 2015, the UCLDC project team released a set of three production services:

- A shared digital asset management system (DAMS) for use by any campus library department
- A metadata harvester that can aggregate and index digital content hosted on any platform.
- A user-focused public interface (Calisphere) that provides seamless access across the entire aggregation, plus an application programming interface that allows for customization and "remix."

As described below, all three of these services are now heavily in use, resulting in broad, aggregated access to an unprecedented number of the University's most unique and remarkable research collections.

The Solution: A Modular Service Stack

The three services released as part of this project together form a pipeline for building and exposing digital collections. But the services also stand alone. The project team deliberately designed the platform to be flexible: previous collaborative digital collection building efforts had failed, in part, because they presented tightly bound services and thus did not offer the campuses, with their distinct priorities, resources, and levels of technical capacity, an opportunity to pick and choose the technologies they needed. The UCLDC project instead created a platform that is modular, allowing library departments to use just the services that complement their own objectives.

Service 1: Shared DAMS

Library collections are large and complicated, and they require systems capable of storing and managing hundreds of thousands—if not millions—of digital files (in a variety of formats), derivatives, and associated metadata. Prior to the completion of the UCLDC project, UC Libraries were on their own when it came to building and/or licensing these systems. And they met with varying levels of success. Although a few had managed to implement sufficiently robust systems, most of the libraries struggled with products that did not fully meet their needs and therefore stymied their digital collections ambitions. In 2009, a task force recommended that the UC Libraries "implement a coordinated, system-wide DAMS solution immediately," emphasizing both the urgency of the need and the cost savings and efficiencies to be gained by a central solution.

The UCLDC project team spent two years installing and customizing Nuxeo, an enterprise management system with an open-source code base. Nuxeo is hosted centrally at the California Digital Library, and any UC campus library can use it through Shibboleth authentication. The UCLDC project team used a strategic, "borrow before build" approach to development. First, the team employed as many Nuxeo features "out of the box" as possible. Next, it utilized plug-ins offered by the product, for example Amazon S3 integration. Finally, the UCLDC team performed selected development work to create features not globally supported by the vendor, such as a bulk file upload client. This tiered approach saved precious development resources for only the most niche, library-specific requirements and allowed the team to deliver two other full-fledged services on the same ambitious timeline.

Service 2: Metadata harvest

The shared DAMS provides one option for campus libraries to store and manage their digital collections. But a few UC Libraries departments are keeping or building their own systems; others are still transitioning to the shared solution; and all of them have legacy content dispersed across a variety of platforms and websites. Prior to this project, these vast resources were "siloed," meaning it was impossible for researchers to locate and search across all of UC's—and in some cases even all of a given library's—collections.

A second major goal of the UCLDC project was, therefore, to develop a way to aggregate all of the unique digital content from across the system. This goal was achieved through the creation of a metadata harvest, based on code from the Digital Public Library of America. Metadata and, sometimes, actual content files from a variety of sources are gathered, stored in a common index, and remediated and normalized for search and discovery.

The harvest has proven to be a particularly scalable aspect of the UCLDC platform. CDL has since opened this service to non-UC institutions throughout California, who also have struggled to gain broad visibility for siloed and under-supported digital content. Not only does this effort align with the public service mission of the university, but it makes for a stronger aggregation with a higher research value: the more collections, the more connections.

Service 3: Public interface

Since the ultimate purpose of the UCLDC project was end-user access, it was imperative that the team create an accessible, engaging public interface through which people could discover UC's archival treasures. An existing website provided an ideal starting point for this aspect of the project. Calisphere was released by the CDL in 2006

with the express goal of making university-level resources available to and usable by non-university audiences. By all accounts, it succeeded. But it needed a technical and visual refresh to serve modern-day users.

The UCLDC project team thus built a new and improved Calisphere website for UC's digital collections (see https://calisphere.org). Drawing from user surveys and testing, analytics, and input from librarians, the team kept the features that had proven successful and incorporated new design ideas. Calisphere now boasts a clean, modern look; lightning-quick retrieval of its hundreds of thousands of items; and helpful new features like filters, library landing pages, and a download button. It is also fully responsive on all devices.

Supporting Campus-Led Development

The UCLDC platform provides not only a set of services to the UC Libraries, but also opportunities for their own development projects. All three of the "layers" of the stack, described above, can be built upon in various ways. The following are some examples of how campus libraries are already taking advantage of such opportunities, and are themselves a part of the ongoing effort to support, improve, and refine the digital collections "pipeline":

- UC Santa Cruz used the Nuxeo API to develop a plug-in for the exhibition system Omeka; this allows campuses to pull objects from the Shared DAMS into custom exhibitions.

- UC Riverside used the Calisphere API to integrate its collections into a one-search feature on its forthcoming new library website.

- UC Irvine is undertaking a project, funded by the National Endowment for the Humanities, to pilot the use of the DAMS as a back-end infrastructure to an experimental linked data application.

- UC Merced is designing and developing a three-dimensional object viewer, based on the Calisphere code base, that will be integrated into the full Calisphere site and made available to all institutions.

Outcomes and Impact

The UCLDC project has revitalized digital collections programs at the UC Libraries. The numbers provide one measure of success: since the project conclusion in September 2015, four campus libraries have become highly active users of the Nuxeo Shared DAMS, and another two campuses are currently in testing. Just in these first six months, the libraries have used the DAMS to manage 68 collections, and to subsequently publish a total of 27 of them to Calisphere--18 of which have *never before been available online.* They include, for example, UC Merced's Dunya Ramicova collection, one of the few complete collections of costume designs by a single designer held by an academic institution; and UC Riverside's extremely rare glass plate negatives of the Mexican Revolution.
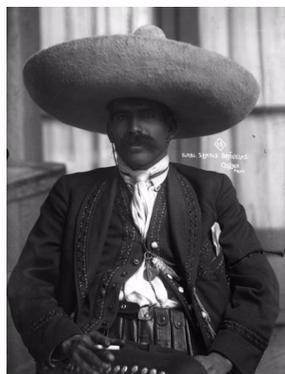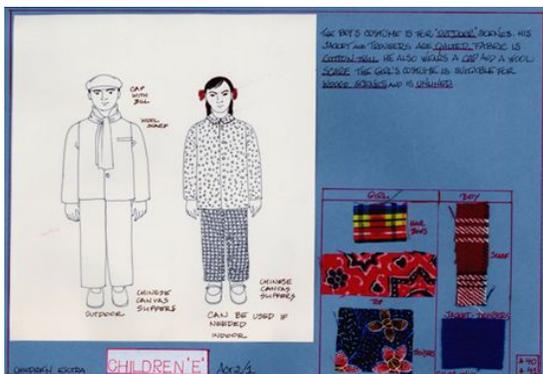
L-R: A costume design from the Dunya Ramicova collection and a photograph from the Sabino Osuna collection of images of the Mexican Revolution.

3

But campus staff provide the most compelling evidence of the impact of this project. In their own words:

"Prior to the launch of UCLDC, the UC Merced Library did not have a system in place for managing digitized special collections or other digital assets we collect [...] We needed to have an easy-to-use, cost-effective system that would handle the variety of formats and complex objects in our collections. Making our content available in Calisphere, where it can be discovered alongside other rich cultural and historical content, rather than in silos, was important to us. [...]The UCLDC suite addresses that full range of needs—it allows us to showcase the valuable collections for which we are stewards, with the aim of supporting the University's three-fold mission."

- Emily Lin, Head, Digital Assets, UC Merced Library

"The UCLDC has played an instrumental role in helping the UCR Library build and provide access to our digital collections. From the ability to effectively manage our digital assets in Nuxeo to delivering this content online through Calisphere, the UCLDC has provided us with the necessary tools to build effective and accessible digital collections for public consumption. The framework has proven essential to the UCR Library as we strive to provide greater online access to our holdings."

- Eric Milenkiewicz, Manuscript Curator, UC Riverside Library

Meanwhile, the UCLDC project has also had an impact on its ultimate audience: researchers. For one, the Calisphere website now provides one-stop public access to an unprecedented number of unique resources, and it is growing faster than ever. When the new Calisphere site was released in September, it immediately contained 400,000 objects or 70% more content than in its previous iteration (due entirely to the scalable harvest infrastructure developed). The aggregation has continued to grow quickly in the months since, notably with the participation of new and previously inactive institutions. It now contains over half a million objects, and the rapid pace of growth is on track to continue.

Preliminary data suggests that the harvest will greatly raise the visibility of harvested collections. Anecdotally, the project team has heard that the usage of locally-hosted collections at the libraries has gone up considerably since the harvest of those collections, suggesting that Calisphere is connecting people with materials that were previously hidden in the more obscure corners of the web.

California State Librarian Greg Lucas sums up the excitement around Calisphere from the end-user perspective:

"Calisphere is one of those resources that librarians, researchers and the global public have been waiting their whole lives for without even knowing it. Not only has it brought together an unparalleled array of primary source material, it has done so in a clean, easily-searchable interface that everyone can use. Calisphere's redesigned site is, in one word, awesome."

- Greg Lucas, California State Librarian

4

UCLDC Project Participants

Project Implementation Team (CDL and UCB)

The project team included staff at the [California Digital Library (CDL) at UCOP](#) and [Research IT at UC Berkeley](#).

- Catherine Mitchell (CDL), Project Director
- Sherri Berger (CDL), Project Manager
- Brian Tingle (CDL), Technical Lead
- Adrian Turner (CDL), Ingest and Operations Manager
- Gregory Shapiro (CDL), User Experience Designer
- Amy Wieliczka (UCB), User Interface Developer
- Joel Hagedorn (CDL), User Interface Developer
- Mark Redar (CDL), Ingest and Operations Developer
- Barbara Hui (CDL), Ingest and Operations Developer
- Yuteh Cheng and Lam Voong (UCB), Ingest Specialists
- Patrick Schmitz, Richard Millet, and Chris Hoffman (UCB), DAMS and Ingest Consultants
- Felicia Poe, Jane Lee, and Eric Satzman (CDL), User Interface Design and Development Consultants

Project Collaborators (all campuses)

The team worked closely with staff across the UC Library system to refine requirements and test and provide feedback on all aspects of the product. Input was formally coordinated through the following two groups, but many additional individuals throughout the UC Libraries contributed their time and expertise to the project.

"Throughout the UCLDC project, the staff at CDL ran a highly collaborative endeavor that sought to seek and implement input from stakeholders across the system."

- Roger Smith, Director, Digital Library Development Program, UC San Diego

UC Libraries Strategic Action Group:

- Lynne Grigsby (UCB)
- Xiaoli Li (UCD)
- Kristine Ferry (UCI)
- Adrian Petrisor (UCI)
- Todd Grappone (UCLA)
- Sara Davidson (UCM)
- Eric Milenkiewicz (UCR)
- Catherine Friedman (UCSD)

- Catherine Friedman (UCSD)
- Susan Boone (UCSF)
- Marti Kallal (UCSB)
- Robin Chandler (UCSC)
- Sue Perry (UCSC)
- Patti Martin (CDL)

Project Stakeholder Group:

- Mary Elings (UCB)
- Sara Gunasekara (UCD)
- Audra Eagle Yun (UCI)
- Stephen Davison (UCLA)
- Emily Lin (UCM)
- Eric Milenkiewicz (UCR)
- Roger Smith (UCSD)
- Polina Ilieva (UCSF)
- Chrissy Rissmeyer (UCSB)
- Sue Perry (UCSC)