

Digital Archives Repository

Project Briefing & Discussion of Summer 2022 Pilot

Prepared by the [CSU Digital Repositories Steering Committee](#)

<i>Introduction</i>	2
Purpose of this document	2
Background information	2
<i>Goals and objectives</i>	2
<i>Project outline</i>	3
Interested campuses.....	3
Pilot.....	4
Next steps.....	4
<i>Resources</i>	5
Costs & budget	5
Staffing.....	5
Governance	6
<i>Timeline</i>	6
<i>Important factors to consider</i>	6
Constraints.....	6
Risks	7
Critical Project Barriers	8
<i>Communications Plan</i>	8

Introduction

Purpose of this document

The Digital Archives Project Briefing & Discussion of Summer 2022 Pilot Document defines the following:

- Project goals and objectives
- Project outline
- Costs, budget, staffing, and timeline
- Constraints and risks

This document is intended to clarify next steps for this project, which will focus on a pilot phase to be carried out over the summer of 2022. The findings of that pilot will inform future planning efforts and an anticipated timeline for implementation of a systemwide digital asset management system.

Background information

For the [past several years](#), the CSU libraries have been engaged in various efforts to upgrade the systems and infrastructure they use to manage digital assets. This includes institutional repositories to archive and promote the research output of each campus, as well as digital archives to manage historical photos, documents, and audio/video recordings.

Nineteen campuses currently use ScholarWorks for their institutional repository. In 2019, the Chancellor's Office implemented a new, centrally hosted version of ScholarWorks based on the Samvera framework and completed the last major data migration from our older DSpace repository to the new system in 2021. Today, ScholarWorks is a bustling institutional repository, with new submissions added daily. It currently houses over 64,000 records with 82,000 files, totaling (with backups) over 4TB of data.

The CSU libraries utilize a more diverse array of systems for managing their digital archives. While five campuses also use the centrally hosted DSpace for this purpose, the remaining campuses use locally supported systems: eight are on CONTENTdm, two use Islandora, two use Digital Commons, and five campuses use either Luna, Omeka, iBase, or another locally developed solution. Additionally, some campuses have digital assets stored on hard drives outside of any system. Although it has been difficult to gather statistics on the size of all these collections, we estimate it likely totals well over 20TB.

Despite this diversity of digital archives systems, two common factors stand out: virtually all campuses are using aging versions of these systems for which there is little support or new development, and/or are using systems that are not designed to address the complete needs of a digital archive (e.g., DSpace and Digital Commons).

Goals and objectives

The technical goals for this project are to implement a single, shared, centrally hosted digital archives system to provide:

- **Core functionality** for describing, uploading, and curating digital assets, including the ability to perform operations in batch.

- **Increased access and visibility** of our digital collections by housing (or harvesting) digital assets from multiple campuses into a single system, enabling cross-references between similar collections held at different libraries.
- **Improved search functionality and user experience**, especially given that many of our current systems have limited search capabilities and outdated interfaces.
- **Improved preservation** through Amazon Glacier or another commercial preservation service.
- **IIIF viewer** to provide support for streaming audio, video, and images, with zooming and panning capabilities.

The business objectives for this project are to:

- **Keep costs low** by implementing the new digital archives using the open source Samvera framework, especially since many campuses lack funds for, and agreement upon, a single commercial alternative.
- **Utilize existing expertise** by implementing the new system using the established technology infrastructure already supporting ScholarWorks.
- **Systematize workflows** and streamline processes by combining expertise from the individual campuses. Shared knowledge moves this project forward.

Project outline

Interested campuses

The Digital Archives Working Group (DAWG) conducted a self-evaluation in 2021 that identified six campuses that are currently ready and willing to migrate to a centrally hosted digital archives:

- Channel Islands
- East Bay
- Fullerton
- Northridge
- San Francisco
- San Marcos

Five other campuses said they might participate:

- Los Angeles
- Moss Landing
- Pomona
- Sacramento
- San Bernardino

All campuses expressed an interest in at least having their content harvested into the digital archives for discovery. From these campuses we will solicit volunteers – which we will term the **vanguard campuses** – to play an active role in the initial phases of the project.

Pilot

While the CSU Libraries have already used the Samvera framework to successfully build a shared institutional repository, the sheer size and diversity of assets that would be housed in a similarly constituted digital archives pose a unique set of challenges. For that reason, the project will begin with a pilot phase to initially design and road test a shared system and inform the requirements and timeline for full implementation across all interested campuses.

The pilot phase of the project will include a small, representative sample of collections to be identified by the vanguard campuses. The collections should come from CONTENTdm, DSpace, and possibly one other system. This will allow the Chancellor's Office to work out system configurations and migration scripts, and to identify areas of needed development.

In consultation with project managers from the vanguard campuses, the Metadata Working Group (MWG) will perform a metadata assessment to identify all schemas, controlled vocabularies, and data types used across the sample collections. Based on this assessment, the MWG will develop a preliminary Metadata Application Profile (MAP) with crosswalks for each collection. Finally, prior to migration, the group will assess the quality and consistency of the metadata and identify areas for immediate remediation and clean-up.

The vanguard campuses will evaluate an initial test system throughout the pilot. Barring any major issues surfaced during that evaluation, this phase of the project will end with the initial implementation of a production system (including authentication) for any vanguard campus that wishes to begin using it.

Next steps

Upon successful completion of the pilot, we will have a clearer understanding of the requirements and timelines for a full implementation of a shared digital archives. Subsequent phases and timelines will be detailed in a future report, but will include:

- Migrating remaining DSpace content
- Migrating remaining vanguard campuses using other systems
- Migrating interested, non-vanguard campuses
- Quality control and assessment
- Sunsetting legacy systems

Like the monetary costs mentioned below, the timeline for these next steps will vary greatly depending on the number of campuses who wish to participate and the size of their collections – which won't be known until after the pilot phase.

Resources

Costs & budget

Currently, the Chancellor's pays about \$24k per year in server and storage costs (using Amazon Web Services) to support ScholarWorks and its attending infrastructure. If we build a digital archives system on that existing infrastructure, we estimate that AWS cost will rise about \$2,250 per TB per year. With 10 TB of data, the total infrastructure costs would rise to \$45k per year.

Additionally, the Chancellor's Office pays \$24k per year to Notch8 for maintenance and development support. We anticipate additional contractor costs to support the new digital archives will be about \$10k per year, for a total support cost of \$34k per year.

Since there are no licensing costs to use the Samvera framework, the total monetary cost of the new digital archives will largely be a factor of how many campuses participate and the size of their collections. The Chancellor's Office can likely cover most, if not all, of these costs for the foreseeable future, but once we exceed 10 TB of data, increasing costs may need to be borne by campuses.

Staffing

Chancellor's Office

The Chancellor's Office currently has two in-house positions and a contract with Samvera contractors Notch8 to support the digital archives.

Director

David Walker, Director of Systemwide Digital Library Services, will oversee the migration and implementation of the new digital archives, serving as both project manager and as the lead developer and user interface designer for the system. Long-term, this position will also oversee the maintenance and support of the system.

Programmer Analyst

Bryan Tu, Programmer Analyst within Academic Technology Services, will do most of the configuration and initial development of the digital archives.

Notch8 contractors

Notch8 will provide consultation, infrastructure and maintenance support, and new feature development for the new digital archives as part of their overall support of the Chancellor's Office's repository infrastructure.

Individual campuses

As with ScholarWorks, participating campuses would need to support this project by identifying campus project managers and ensuring regular representation on related committees and working groups.

Project Managers

For the pilot phase, project managers would need to commit to active involvement in the platform testing and metadata work outlined above. While the Metadata Working Group is prepared to provide direction

and support on this work, in coordination with the Digital Archives Working Group, the former's members are simultaneously supporting ScholarWorks and do not have capacity to do both without additional help.

Governance

Effective management and shared governance of the digital archives will be an essential component of the project's success. The Digital Repositories Committee (DRC) will provide overarching governance of the system, addressing specific areas of need through its two working groups, MWG and DAWG, as well as any task forces it creates.

Timeline

March 2022	Identify vanguard campuses and pilot collections
April 2022	Begin pilot
August 2022	Finish pilot project

Important factors to consider

Constraints

The limited staffing and budget for digital repositories projects in the CSU, including both ScholarWorks and the digital asset management system, remain the most pressing constraints. While the Samvera-based ScholarWorks is live and being used by many campuses, by virtue of being an open-source solution that was not designed for consortial use, development and maintenance is ongoing. The success of that work is dependent on the active engagement of campus managers for each campus as well as ongoing support from the Metadata Working Group, which is continuing to work towards standardizing metadata fields and coordinating metadata cleanup for ScholarWorks.

What this work has made clear is that most individual campuses need additional training and guidance to manage their institutional repository assets, which raises concerns as to whether individual campuses can maintain that workload while also moving forward with a systemwide digital asset management system. In addition to the Metadata Working Group, a Work Forms Task Force was also formed in the fall of 2021 to address the need for more user-friendly and responsive submission forms for ScholarWorks. Led by the Director of Systemwide Digital Library Services, this group is also composed of volunteers, two of whom are also serving on the Metadata Working Group, further demonstrating the challenges this project has encountered from a human resources perspective.

In the best of cases, building a digital asset management system in the absence of additional staffing and budget will make our progress very slow. As the [SWAT report in 2019](#) found:

"The most successful repositories in our sample tended to:

- Spend a significant amount of money on salaries or fees for support staffing
- Bolster funding through grants and consortial costs and appropriate membership fees.

- Prioritize services, especially the CDL, Oregon Digital, TDL and WRLC.
The platform is merely the vehicle to fulfill the current policy and governance charges.”

We are working toward a California Digital Library style repository, but without the staffing levels or budget of the CDL. Additional staffing and more funding are needed to speed up migrations and development as well as allow the project to move forward during any disruptions.

Risks

The following events pose risks to the project:

1. CO staffing turnover

Risk level: High

Likelihood: Somewhat likely

Impact: The project would likely slow down significantly, especially if both positions turn over at the same time, as that would greatly impact continuity.

Mitigation strategy: The Chancellor’s Office will rehire and train new staff, and may be able to put more money into contractors during the transition.

2. Reduction in central funding

Risk level: High

Likelihood: Low

Impact: The project as defined here would be untenable if costs cannot be fully covered.

Mitigation strategy: Campuses could pool money to cover AWS costs, perhaps even moving the system fully to Notch8 for hosting. Alternately, limited funding could support a temporary time frame to export content to the campuses.

3. Staffing turnover at individual campuses

Risk level: Medium

Likelihood: Medium

Impact: Without a local staff person to oversee the export of content and mapping of metadata to the digital archives, a campus will see significant delays in migrating to the new system.

Mitigation strategy: Campuses can ensure additional library staff have knowledge of local systems to participate in the migration.

4. Problems uncovered in the pilot

Risk level: Medium

Likelihood: Medium

Impact: If significant new features are needed before some campuses will join the new digital archives, project timelines for subsequent phases of the implementation will be adversely affected.

Mitigation strategy: DAWG is currently completing its analysis of Samvera to identify any areas that may need new development.

5. Turnover on the working groups

Risk level: Low

Likelihood: High

Impact: In areas where the project is impacted by governance – such as in defining and standardizing metadata or reviewing new features and functionality – progress would slow until additional volunteers can be located.

Mitigation strategy: Question for COLD: How do we get more participation from the campuses that haven't been participating?

Critical Project Barriers

Unlike risks, critical project barriers are insurmountable issues that can be destructive to the project's initiative. In this project, the following are possible critical barriers:

- Major reduction in state funding to the CSU.
- Removal of central project funding and positions.

Should any of these events occur, the project plan would become invalid.

Communications Plan

Disseminating knowledge about the project is essential to its success. The Digital Repositories community will continue to use the established communication channels for this work:

- Monthly “open forum” meeting
- Bi-monthly email update
- Slack channel
- Digital Repositories listserv
- Annual meeting
- Digital Repositories Steering Committee Annual Report to COLD