California State University

CSU Libraries DAMS Report

Requirements for Implementing Shared Digital Library Services

Prepared by the Digital Archives Interest Group

FINAL REPORT
August 9, 2019

Stephen Kutay, California State University, Northridge
Alyssa Loera, Cal Poly Pomona
Kevin Cloud, California State University, Office of the Chancellor

Table of Contents

Executive Summary

In March of 2018, a volunteer working group comprised of archivists and digital collections supervisors convened to extend the work of the SWAT ScholarWorks team regarding the use of Samvera and associated technologies used for digital asset management of institutional repositories. The charge of the Digital Archives Interest Group (DAIG) was approved by STIM with the objective to “explore the possibility of CO hosting of ArchivesSpace, the current industry standard collection management tool, and to develop a digital asset management system based on the Scholarworks platform already in use to support the creation, discovery and preservation of digital archives.”

To meet its charge, the group divided into three functional areas to study 1) the potential of a unified instance of ArchiveSpace to develop and serve finding aids, (later suspended due to lack of resources), 2) the potential of the Samvera technology stack to provide digital asset management of digital collections resources across the CSU, and 3) a metadata team designated to begin development on a data dictionary to build common data models represented in digital collections of CSU libraries.

In order to provide data to analyze digital asset management needs specific to digital collections, a survey was distributed in November 2018 and compiled into the *Digital Collections and Archives Survey Report* in January 2019 (Appendix B). Major takeaways include the perceived advantages and disadvantages below.

Popular perceived benefits of adopting a shared digital asset management system:

- Searching and collaborating across all CSU collections
- Improved discovery through harvests
- Improved interface and object viewers
- Some preservation support
- Streaming support
- Improved tech support and workflows
- Cost savings

Popular perceived detriments of adopting a shared digital asset management system:

- Loss of administration/management control
- Loss of system and user customization/local branding
- Challenges to local workflow needs
- Bottlenecks (in system performance, services, and support)
- Reaching consensus between campuses regarding development/priorities

The functional requirements and recommendations offered in this report are intended to realize perceived advantages and expressed needs and mitigate perceived disadvantages associated with a unified digital asset management system for digital collections of the CSU.

Summary Recommendations

Data gathered and analyzed in this report were used to develop the following five recommendations intended to effectively integrate digital collections into the shared DAMS of the CSU Libraries. See *Recommendations*, p. 23 for a more detailed analysis.

1. **Apply the functional requirements as presented in this report as a map for guiding the implementation of a shared DAMS that effectively serves the needs of digital collections production and management within the CSU**. The functional requirements reveal priorities of the survey respondents as expressed in the data, as well as illustrate other functions and configurations necessary to ensure comprehensive digital asset management. Each functional requirement is assigned “R” (required) for needs designated as most important in the survey, or “P” (preferred) to acknowledge that not all priorities, though desired, can be guaranteed.
2. **Include within the official DAMS governance body (i.e. Digital Repositories Steering Committee), an advisory group of representatives comprised of digital services, collections, initiatives, or other personnel from participating campuses to advocate on behalf of stakeholders and users of both institutional repositories and digital collections (Digital Archives) respectively.** Such a group should also be responsible for creating bylaws and/or procedures that support equitable voting and decision-making across participating IRs (ScholarWorks) and digital collections personnel.
3. **Formalize a *working group* of metadata personnel drawn from participating campuses to lead the continued development of shared metadata practices, formatting, data models and assessments.** This recommendation supports and extends the valuable work already completed by Chancellor’s Office (CO) staff and members of the Metadata/Linked Data Interest Group. It also recognizes that, as a fundamental component of digital asset management, metadata applications and iterative assessment are critical to establishing both quality and consistency in our quest to make accessible the objects contained within the DAMS, as well as serve the needs of *all* its users.
4. **Analyze the needs, feasibility, and costs associated with adding additional labor necessary to assist CO staff in the management and development of a shared DAMS for digital collections and IRs**. The success of a shared repository will depend in part on increased commitments of labor and/or expertise potentially through the addition of support staff or third party services. A group such as a task force charged with assessing needs and additional costs associated with a shared DAMS could well serve its continued development and overall management.
5. **Assess the willingness and capacity of campuses to help design and program features, within (or external to) the DAMS API that will improve management of the system and user experience**. By definition, open source technologies excel when users make and share their improvements to the platform or application. A task force or other small group assembled from existing Interest Groups could survey campuses for available voluntary talent in these areas.

Report Objectives

The intention of this document is to assess the digital library needs of the 23 campuses (as determined by the SWAT team survey) and translate those needs into technical requirements for a shared digital asset management system. The provided baseline analysis of survey data combined with guidance from multiple DAMS evaluations and environmental scans will help to define an acceptable and achievable scope of services. No one set of recommendations can comprehensively represent all the diverse needs of multiple organizations in a single platform. For this reason, alternative third party integrations are suggested where appropriate. Finally, this report provides a resource for assisting with the projection of costs of implementation and maintenance.

A glossary (APPENDIX A) is provided to describe the usage of some terms used in this report.

Background

The Scholarworks Initiative has existed in some form for approximately twelve years, predominantly as a mechanism for CSU-wide institutional repository collaboration. With the installation of DSpace instances at multiple campuses (all hosted through the Chancellor's Office), each campus developed workflows, policies, and branding independent of each other. In some cases these practices were only applied to item types that are typically considered institutional repository materials, in other cases this also applied to item types typically hosted through a digital library platform.

In October of 2018 the Council of Library Deans (COLD) voted to charge a SWAT Team (a task force consisting of individuals from the 23 campuses within the CSU) with the development of a cost-benefit analysis regarding the feasibility of moving from multiple, per campus, DSpace instances into a single Samvera/Hyrax back-end and public facing portal. This team sent out two surveys to all campuses to gauge the foundational and desired requirements of an institutional repository, and also for a digital library. From that cost-benefit analysis, three recommendations were put forward to COLD:

1. Implement a single-instance institutional repository. Although there are clearly some trade-offs with this approach, it allows us to more optimally use our limited central resources over the medium- and long-term.
2. Create a more formal governance structure for ScholarWorks. The current COLD committees (STIM and ScholCom) and the community of practice around ScholarWorks have overlapping or uncertain authority with regard to the IR. COLD should create a new governance structure (more like ULMS) to make the shared decisions (such as scope and metadata) needed with a single IR. The successful consortia we surveyed all have clear governance structures.
3. Approve a policy regarding scope of content for the IR. The SWAT team will be submitting a proposed policy to COLD separately from this report. We recommend that

institutional archives content currently in ScholarWorks that falls outside of this scope remain in DSpace until a decision is made around a replacement digital library platform.

Recommendations one and three were approved by COLD on March 1st, 2019. The next iteration of the SWAT team will provide more detail about recommendation two to COLD before moving forward with a revised governance structure.

Recommendation three has evolved into a draft policy regarding how item types within an institutional repository compares with those in a digital collections repository that hosts university content. Such institutionally-created items considered for transfer from IRs to digital collections include the following:

Administrative documents
Meeting Agendas
Newsletters
Ephemera (flyers, programs)
Student publications (magazines, newspapers, newsletters)
Student organizations' output (publications, events, images)
Photographs
Fine arts collections
Multi-page documents (letters, reports, newsletters)
Sound recordings
Newspapers
Transcripts/interviews
Posters
Architectural drawings
Maps
Aerial Photographs
Realia
Fashion plates
Musical scores
Oral histories, including audio and transcripts - unless they’re part of a student/faculty research project (then IR)
University Concerts
Recitals
Museum Collections/Catalog or Citation Data

Whereas all other items will exist within the single portal IR. Although this scope policy is only a draft, it still will inform the baseline needs of a single digital library platform or stack, one that is separate from the current Samvera beta migrations. The ideal specifications outlined by this report will accommodate these item types.

Methods and Limitations

This report of the functional requirements for Samvera implementation for digital collections uses multiple sources to support its conclusions and recommendations. Central to these is the *Digital Collections and Archives Survey Report* (Appendix B) compiled by the Scholarworks SWAT Team in January 2019. This report gathers data received from the November 2018 survey that illuminates existing budgets, staffing, and priorities toward a system-wide instance of a Samvera digital asset management system (DAMS) across the twenty-three CSU campuses.

Limitations are represented by three primary factors: 1) Data received regarding preferred functionality of a shared Samvera DAMS did not measure satisfaction for currently deployed systems; 2) Data was not collected on the level of digital collections activity (vigor) per campus, and; 3) Challenges created by distance and time prevented more nuanced data gathering methods such as focus groups or interviews. This report therefore also considers three environmental scans (see Related Studies) on the subject produced by other academic institutions and consortia. (Appendix C)

Related Studies

Several institutions nationally have conducted environmental scans for identifying motivations for migrating to new DAMS as well as exploring solutions for shared architectures within digital collections and exhibits. In conjunction with the results of our survey, these studies provide additional knowledge and perspectives toward the consideration of shared digital library services.

In 2015, Ayla Stein and Santi Thompson produced “Taking Control: Identifying Motivations for Migrating Library Digital Asset Management Systems.” The authors surveyed libraries, archives, and museums around the world who had migrated their existing digital libraries to new systems. Of the 49 respondents, 61% were from academic institutions. The survey revealed the top areas of concern for migrating to a new DAMS were: 1) Content Management; 2) Metadata; 3) Extensibility; 4) Preservation, and; 5) User Interface Customization.

Within content management, respondents most frequently cited as a concern poor system performance as a result of increasing numbers of digital objects, and the ability to perform batch uploads. Robust support of various file formats also ranked highly.

Given that some proprietary systems offer limited metadata schema support, respondents desired the ability to apply a greater variety of schemas. Among the most popular in order of preference were Dublin Core, MODS, and EAD. Respondents also indicated a desire for a

variety of standards such as METS and PREMIS in addition to linked data technologies RDF/XML, and JSON.

Regarding extensibility, respondents most desired local control for creating modules, plugins and similar functionality in a system with an application programming interface (API), and shared code. However, in recognizing that some organizations lack necessary resources for such customizations, there was also adequate desire for governing institutions to lead programming efforts.

Preservation of digital assets were of a concern of respondents focused on tasks that could be delivered in a system to ensure the integrity of their files, such as checksum verification. Respondents also indicated a preference for producing Archival Information Packages (AIP).

Respondents desired local control over user interface customization in order to match local needs, as well as utilize branding familiar to such users. Responsive design that enabled use on mobile technologies also was designated as important to respondents.

In addition to the above priorities, dedicated and prompt technical support was also highly desired among respondents.

Having identified the above motivations for migrating to new DAMS, Stein and Thompson also surveyed respondents' preference for moving to proprietary, open source, or home-grown systems. At the time of the survey, over half the respondents originally used proprietary DAMS. However, nearly two-thirds selected open source systems to migrate their collections, suggesting a trend among libraries, archives, and museums.

In 2016, University of California (UC) reevaluated their existing shared DAMS against a Fedora/Hydra technology stack in *UC Libraries DAMS Technology Report: Assessment of a Long-Term Solution for the UC Libraries Systemwide DAMS* (Smart et al., 2016). At the time of this report, the UC shared digital collections database, known as Calisphere, was developed and maintained by the California Digital Library utilizing Nuxeo, an open source digital asset management system to host digital collections from University of California and many other libraries throughout the state.

The report established the following "key themes" that informed further recommendations. First, the Fedora/Hydra technology stack (now Samvera) served as a compelling choice for more fully developing the shared DAMS. Second, UC libraries expressed interest in co-developing or implementing technologies and other services associated with the digital collections infrastructure. Third, in order to play a role in service development, the multiple campuses of the UC would have difficulty signing onto a collective agreement as a contributor to the development of the Fedora/Hydra platform under DuraSpace. And four, the existing DAMS must remain fully functional throughout the development of any new system.

Stemming from the fourth theme, the committee formally recommended that the existing Nuxeo system be used to provide digital collections services for the near- and mid-term future. The second recommendation was for a collaborative model of shared DAMS infrastructure co-development between the ten UC campuses and CDL, and to investigate how the use of their agile development program could be leveraged for this purpose. The third recommendation asked CDL to negotiate with DuraSpace a license agreement with respect to copyrightable contributions made by the UC co-development team. A fourth recommendation urged a test implementation of the turnkey version, Hydra-in-a box (released as Hyku/Samvera in 2017), in order to begin learning the technology stack system-wide. In its fifth and final recommendation, the report team saw DAMS development as an opportunity to investigate the benefits and costs of linked data integrations. A link to the 2018 resultant report on the UC linked data analysis is also included in Appendix C of this report.

In 2018, Dawn Childress and Pete Broadwell of the UCLA Digital Library Program conducted an environmental scan at the campus level in order to suggest long term strategies for UCLA's Digital Library Services, specifically. This white paper contains user stories, short term and long term recommendations for digital collections, and an evaluation of current digital library systems.

The research in this paper is based on stakeholder interviews, focus groups, and system use observations. The key arguments in the report are that a DAMS should be content-agnostic, interoperable, and flexible. More specifically, the development of a core DAMS should be, "based on the argument that activities around managing, publishing, and preserving assets should be decoupled, yet interoperable" (pg. 8). On top of a core infrastructure, Broadwell and Childress argue the need for interoperability in the form of customizable and interchangeable specialty services. Ultimately, the argument for a long-term transition into Hyrax and the customizable components built into the Samvera system is made.

The report also indicates the abandonment of the Islandora software stack was due to a lack of interoperability with Drupal, and a general lack of support for named use cases. The report cites its reasons to suspend Islandora development in the *UC Libraries DAMS Technology Report* (Appendix C) largely on the basis of UCLA's unique test implementation outside of the larger Islandora community, and that Islandora modules designed for Drupal by the community were not well enough integrated into all of Drupal. Further, any future development would be hampered by the transitions from Islandora 7/Fedora 3 with Drupal 7 to Islandora CLAW/Fedora 4 with Drupal 8 (pp.29-30). Because of the complex nature of developing a new Samvera instance, the short-term goals included the continuance of using the Nuxeo DAMS, with support and infrastructure from the CDL.

Survey Results

Below is a summary of the findings of the Digital Collections and Archives Survey Report created by the ScholarWorks SWAT Team, January 2019. For more information, see APPENDIX B.

Respondent activity

Survey respondents include the twenty-three primary personnel responsible for digital library or digital initiative development across all campuses of the CSU system.

Systems Deployment

Digital collections systems deployed with annual systems costs among campuses (n=23):

1. Campus instance of platform (n=12) - $168,678
2. CO-hosted DSpace platform (n=8) - $64,140
3. Does not have a digital collections service (n=3) - $0

Perceived CSU Benefits

Expressed benefits of a system wide digital collections platform include (in order of preference):

- Search all campus collections
- Built in preservation tools
- Improved discovery through regional/national metadata harvests
- Improved and responsive UI
- (Tied) A/V streaming; Improved object display and functionality (e.g. IIIF); Persistent identifiers; Usage statistics and reports

Perceived Disadvantages

Perceived disadvantages imposed by a system wide digital collections platform include (in order of popularity):

1. Challenges to local administration of site
2. Challenges to UI customization
3. Challenges to custom submission workflows
4. Challenges to custom permission levels
5. Challenges to collections (or landing) pages, exhibit pages, and local search facets

Perceived Campus Benefits

Perceived benefits to individual campuses regarding a system wide digital collections platform include:

- Increased access through intercampus searching and exhibits of related holdings
- Improved tech support through Chancellor's Office
- Improved workflows and management
- Cost savings
- Increased intercampus collaboration
- Improved user functionality
- Preservation tools

Essential Features from Existing Platform

Expressed essential features of a unified digital collections platform that is already established in campus' existing platforms include:

- Robust search capabilities
- Multi-format display functionality (e.g. books, images, sound)
- Anything deemed "Very Important" from the survey

Perceived Detriments

Perceived detriments to a system wide digital collections platform include:

- Loss of administration/management, customization, and control
- Loss of autonomy over metadata, templates, and collection management
- Campus location of physical objects listed in federated search results
- Competing campus priorities skewing performance/functionality priorities
- Reaching Consensus
- Bottlenecks
- Challenges to institutional needs regarding policies workflows, collection structures, metadata, and vocabularies/thesauri.

Functional Priorities

Top priorities for DAMS functionality for digital collections (in order of popularity):

1. Keyword search
2. Full text search

3. Multiple formats and multimedia support
4. Title and Author search
5. Usage statistics/submissions reports
6. Audio and video streaming support
7. Filters and faceted search
8. Preservation

Feature Priorities

Top priorities regarding **features** of a system wide digital collections platform include (in order of popularity):

1. Digital Asset Management
2. Advanced and embedded image viewer
3. Video player
4. Audio player
5. Local branding
6. Custom workflow support
7. Compound objects
8. Preservation

Ancillary Systems Priority

Top priority for ancillary systems:

1. Exhibits

Additional Features

Additional features requested in a system-wide digital collections platform (no preference):

- Preservation functions (fixity/integrity checks, format migration)
- OCR
- Technical metadata (extraction)
- User tagging/description
- IIIF
- Batch/Bulk uploads
- Global editing
- Consistent metadata
- Templates
- Born-digital ingest and preservation
- Rights management

Top Document Types in Existing Digital Collections

Top document types (in order of popularity):

1. Photographs
2. Paper records and publications such as multi-page letters, reports, newsletters
3. Sound recordings
4. Video recordings
5. Newspapers/Transcripts/Interviews (Tie)

Integration rates with external discovery systems

Do you integrate with external discovery systems?

Yes	52.2%
No	47.8%

With whom do you share your metadata? (n=13)

Calisphere/DPLA	53.8%
OAC	23.1%
Primo	15.4%
Internet Archive	7.7%

Discussion/Analysis

This section includes analysis of notable points in the data used to inform priorities in a shared DAMS. Desired functions and features common to most, if not all, asset management products are not discussed in favor of focusing more attention on those that are more unique, impactful, or essential to the design of a shared Samvera implementation. For additional analyses of the survey results see *Digital Collections and Archives Survey Report* (Appendix B).

Current DAMS Implementations

While digital collections support throughout the CSU is in its early stages, trends suggest a substantial need and benefit for shared DAMS support for digital collections. Of the twenty campuses with digital collections programs, sixty (60) percent are leveraging DSpace, the Chancellor's Office DAMS used for ScholarWorks, to serve users of archival resources most typical of digital collections. Further, those using CO support spend nothing on licensing and maintenance compared to $168,678 spent by the eight institutions committed to locally implemented DAMS solutions.

Perceived Benefits of a Shared DAMS

Perceived benefits as a result of a shared DAMS for digital collections are numerous. For users, the ability to search all campus collections was cited by seventy-eight (78) percent of the respondents. However, the majority of benefits are likely to come from CO support resulting in cost savings, improved workflows, normalized metadata, increased harvests, streaming media support, persistent IDs, and usage reports. Also recognized are functions of the Samvera technology stack that provide improved displays with IIIF viewers and access to preservation tools. Campuses will also benefit from mediated collaborations in the form of exhibits and intercampus grant projects.

Perceived Detriments of a Shared DAMS

Campuses express notable concern over losing control of administering their own webpages and user interface customizations, submission workflows, and user permissions.

Prioritized Functions and Features

Digital Asset Management was identified by ninety-six (96) percent of the respondents as a top priority. Though no specific tasks related to asset management were listed as preferences within the survey, the popular response clearly demonstrates both interest and need across libraries of the CSU.

Respondents also expressed a need for multimedia support with sixty-five (65) and seventy (70) percent of respondents requesting built-in audio and video players respectively. Support for

multimedia streaming was identified as ‘very important’ by seventy-eight (78) percent of the respondents, and ‘somewhat important by an additional thirteen (13) percent of the respondents. While DAMS support viewing multimedia files, streaming media will require additional infrastructure and maintenance to reliably serve a/v files to users without potentially degrading the performance of the overall system. Streaming support, therefore, is considered essential to a shared DAMS.

References to preservation lifecycle management appears multiple times throughout the survey. Specific preservation tasks and tools were not given in the survey, nor offered by respondents, with the exception of fixity checks (checksums captured to ensure the integrity of files within the system), which is provided in a Fedora-based repository. Samvera and its associated technology stack (i.e. Fedora, Solr, and associated middleware and plugins) does not provide comprehensive preservation management services found in dedicated systems such as Archivematica, Rosetta, etc. With that said, preservation management tasks desired by CSU campuses will benefit by shared data models and a Metadata Application Profile (MAP). The data models and MAP provide essential elements and crosswalks that can be utilized in the construction of information packages as part of the OAIS reference model and used alongside various open source software (OSS) used in preservation management workflows. (See Recommendations/Preservation Management.)

Local branding intended to serve institutional identity was selected as ‘very important’ by fifty-seven (57) percent of the respondents, and ‘somewhat important by an additional thirteen (13) percent. While the extent of branding priorities was not measured, it should be considered that minimal provisions of custom color assignments, wordmarks, and logos be available to each campus’ collections pages.

Functional Requirements

The following recommendations represent fundamental service requirements with emphasis on functionality distilled from survey data **(in bold)** and other relevant sources. These functional requirements are intended to provide the basis for evaluating Samvera (and associated technologies) as a shared digital repository platform (referred herein as *system*) among the campuses of the California State University. Each recommendation will be accompanied by one of two priority levels: *required* (R) or *preferred* (P).

Backend Specifications

User Management

The system should supply granular levels of permissions to users of the system in order to meet the needs of flexible workflows. (R)

User roles should include:

- *Super-User* (or administrative equivalent) that has complete configurable access
- *Project Administrator* to provide the highest levels of database maintenance and assign users with subsequent permissions
- *Users* with levels of regulated access to specific features and collections in the system.
- *Public User* accounts to enable personal tagging/organization of online items

Authentication

- While the vast majority of collections will be freely accessible to the public, the system should leverage authentication technologies to restrict applicable collections to an institution's user base as needed. The application of effective authentication technologies, such as Shibboleth, should be used to manage access across variable user groups, thus expanding the overall flexibility of the DAMS to serve a variety of educational purposes otherwise confined by restrictive access relating to embargoes, copyright, etc. (R)

Content types

- **Both born-digital and digitized content will be migrated to IR and digital collections environments. Therefore the DAMS must be format-agnostic to ensure all file types can be managed as needed.** (R)
- In the event an ingested file must be changed, the system should accommodate the swapping of new versions of files. (P)
- **System should effectively serve streaming media content of audio and video formats**. (R)

Object Handling

- The processing of preservation quality formats as access derivatives, such as TIFFs to JPGs, will provide an optional workflow for digital preservation of analog materials as well as valuable automation. (P)
- **Object level rights management and branding should be enabled through the automatic processing of watermarks and/or embedded banners at the time of ingest.** (P)
- **The system should provide automated OCR at ingest for full-text searching of applicable textual materials with typeface.** (R)

Common Data Model

- With a single Hyrax implementation serving up to 23 campuses, a common (shared) data model ensures that all campuses agree upon how content is structured and understood by the system. This allows updates to work equally across all campuses rather than updating multiple Samvera instances with different code to operate with disparate data models for each campus. Using the Resource Description Framework, the model should represent classes (i.e. collections, objects, files) and associated properties (members, relationships). (R)

Metadata

Schema Assignment

- **Descriptive schemas Dublin Core, VRA Core 3.0, PBCore are commonly used for non-bibliographic materials.** (R)
- Bibliographic descriptive schemas such as MODS will add to the extensibility of the platform. (P)

Controlled Vocabularies and Linked Data

- **As a best practice of resource description, the DAMS system must enable the management of controlled vocabularies derived from both standard thesauri and local controlled vocabularies**. (R)
- **Linked data implementation that provides direct access to standard vocabularies will promote consistent descriptive practice between items and collections**. (P)

Remediation/Editing

- The platform should offer object level editing capabilities to facilitate metadata updates to ingested materials. (R)

Find & replace/Fill

- **The platform should offer global metadata editing capabilities within specific fields, such as searching and replacing existing values within a single field, as well as blanket changes to entire field values, such as rights statements.** (R)

Ingest

- **The system should provide batch (or bulk) ingest of multiple items of simple objects of files and metadata**. (R)
- **The system should provide batch (or bulk) ingest of multiple items of compound objects of component files and metadata**. (P)
- The system should provide a user interface for adding individual objects (simple or compound) directly into the system. (P)

Quality Control

- The system should provide quality control functions to administrators for approving, removing, or updating submitted objects within a cue prior to the final stage of ingest, or the building of a collection's index. (P)

Exports

- A collection's full metadata should be exportable for external remediation, preservation, or as downloadable files for public use such as research and machine learning operations. (R)
- Results from a search of a collection's contents should be exportable for external remediation, or entry into other collections within the system. (R)
- Exports should be offered in multiple formats such as delimited text, XML schemas, and METS to provide flexibility when working with databases, or for long-term preservation. (R)
- **System exports should include the extraction of technical metadata of items within the database.** (P)

Back-up/Sustainability

- Redundancy of files and metadata to protect against catastrophic loss. (R)
- Databases must be fully recoverable in the event of catastrophic failure. (R)
- System should be fault tolerant to continue operation during erroneous or compromised performance. (P)

External System Integration (Interoperability)

- **The system must enable the crawling of a collection's metadata and URLs as desired via OAI-PMH to be used within an external discovery systems such as Calisphere and the Digital Public Library of America.** (R)
- Records within the system should be harvestable by the CSU Unified Library Management System (Ex Libris Alma) via an external API or another protocol. (P)

Reports

- **Collection-level and item-level usage statistics and analytic reports will provide valuable data to enable administrators to strategize development and promotion of their collections.** (R)

Preservation Management

It is acknowledged that a digital repository system does not guarantee comprehensive preservation lifecycle management. However, some of the previous actions described in the recommendations above satisfy components of a larger preservation workflow. Fundamental preservation features should include:

- Redundant file storage (see Object Handling) (P)
- Metadata exports in XML/METS (see Exports) (R)
- Integrity checksums (e.g. MD5, SHA256) (P)
- Format validations (e.g. JHOVE) (P)
- Automated reformatting of obsolescent formats (P)
- Ingest of born-digital materials for preservation (P)

Alternatives to built-in preservation management may be handled by each campus individually in the form of multiple open source or subscribed preservation management tools:

- DataAccessioner http://dataaccessioner.org/
- DROID https://digital-preservation.github.io/droid/
- Bagger https://github.com/LibraryOfCongress/bagger
- Bitcurator https://bitcurator.net/
- More at the Library of Congress https://www.loc.gov/preservation/digital/
- More at POWWR Tool Grid https://digitalpowrr.niu.edu/digital-preservation-101/tool-grid/ and Tool Grid v.2 http://www.digipres.org/tools/ubergrid/

Frontend Specifications

Discovery

- Search engine should provide advanced search (multiple boolean and specified field search bars) (R)
- **Search engine should provide, specifically, keyword, author and title search capabilities.** (R)
- **Discovery layer should provide customizable filters and facets** (R)

Delivery Interface

- **PDF viewers should be enabled and effective regarding zooming, downloading, and printing**. (R)

- **IIIF and other embedded viewers will facilitate advanced analysis of image files**. (P)
- **Audiovisual media players should enable basic functions such as display sizing, rewind and fast forward**. (R)

Web pages and customizations

- The system should provide the means for displaying lists of collections (e.g. splash pages) as well as the means for displaying these according to each participating campus. This could be a native function of the software and/or a customizable web-based option. (R)
- Likewise, the system should provide the means for providing textual descriptions (e.g. landing pages) of the collections in order to give users critical context associated with the materials contained therein. This could be a native function of the software and/or a custom web-based option, such as a WYSIWYG editor. (R)
- **The system should provide the means for campus branding of the collections through graphics, wordmarks, and color assignments** (P)
- **The user interface must be responsive according to web standards.** (R)
- **The system must be compliant with the Americans with Disabilities Act (ADA) to ensure accessibility, or is committed to reaching full compliance**. (R)
- With each item created, the system should be able to associate that item with a direct link to the object file that can then be reused in third party tools (e.g. exhibit platforms) for image rendering and object embedding using a stable URL. (R)

Exhibits

- **As an identified priority of the survey respondents, it is requested that the system provide the means for curating exhibits online. One complicating factor is that no mutually agreed upon definition of what constitutes an exhibit online was achieved. Digital repository systems, in general, supply the means to create web pages (see Web Pages and Customizations), however limited, toward an exhibit-like function. In the absence of a dedicated exhibit function, the system should utilize plug-ins designed for this purpose.** (P)

Alternative to the absence of exhibit functions (or plugins) within the system, it is recommended that participating campuses consider the following interoperable third party tools for creating exhibits:

- Spotlight is a free, open source exhibit application that is an extension of the Blacklight discovery platform. It is designed to integrate directly with digital collections items. Since Samvera uses Blacklight and Solr, Spotlight is among the best choices for easily combining existing digital collections resources directly into a Spotlight exhibit. Stanford offers the use of their Spotlight for institutions that wish to avoid local installations and collaborate using materials from the Stanford Digital Repository (SDR). Originally developed by the University of Virginia, Blacklight is now supported by a growing

community of users. A shared instance using existing Samvera technologies would be most desirable.
- From the Alliance for Networking Visual Culture, Scalar offers a media-rich scholarly publishing platform as a free hosted service at the University of Southern California, or as an open source local deployment. Scalar offers both media imports and embedding from other repositories.
- Omeka Classic offers the Exhibit Builder plugin, but requires that selected resources and metadata be ingested into this platform. Omeka is supported by the Corporation for Digital Scholarship, the Roy Rosenzweig Center for History and New Media, and George Mason University, with funding from multiple organizations. Omeka Classic is free, open support (unless a support plan is purchased).

Downloads

- The system should provide full and partial resolution downloads of the materials within the database to enable flexible use in research and course learning objectives. (P)
- The system should enable printing of reasonably renderable file types within the database. (R)

User tags/annotations

- **User functionality is greatly improved by the ability for the public to tag individual items for future use.** (P)
- **Users and custodians of archival materials benefit significantly from the ability of users to annotate. (i.e. provide public descriptions) of the database items therein. (P)**

Social media

- The system should enable sharing through social media by building in links to popular social media sites accessible at the item level through the use of stable URLs and embed codes, or in the least, provide easy access to stable URLs and embed codes. (P)

Recommendations

Data gathered and analyzed in this report were used to develop the following five recommendations intended to effectively integrate digital collections into the shared DAMS of the CSU Libraries.

1. **Apply the functional requirements as presented in this report as a map for guiding the implementation of a shared DAMS that effectively serves the needs of digital collections production and management within the CSU**. The functional requirements reveal priorities of the survey respondents as expressed in the data, as well as illustrate other functions and configurations necessary to ensure comprehensive digital asset management. In order to place the systems functionality in the context of the data and the priorities outlined, the above functional requirements are assigned “R” (required) for needs designated as most important in the survey, or “P” (preferred) to acknowledge that not all priorities, though desired, can be guaranteed. When taken together, the requirements represent a baseline set of functions common to, and expected of the DAMS. In cases where the DAMS cannot satisfactorily provide functions to the degree requested by respondents, alternate ancillary open source alternatives are recommended as supplemental technologies to the shared DAMS.

2. **Include within the official DAMS governance body (i.e. Digital Repositories Steering Committee), an advisory group of representatives comprised of digital services, collections, initiatives, or other personnel from participating campuses to advocate on behalf of stakeholders and users of both institutional repositories and digital collections (Digital Archives) respectively.** Such a group should also be responsible for creating bylaws and/or procedures that support equitable voting and decision-making across participating IRs (ScholarWorks) and digital collections personnel.

   The need to establish fair and representative governance of the system arises from multiple factors.

   First, the history of the development of the shared DAMS and the initial inquiry into the Samvera repository has been borne out of the desire to replace DSpace as the asset management system used for CSU IRs (ScholarWorks). Early development for this purpose provides a structural advantage to institutional repositories that have significantly different workflows and users from digital collections.

   Second, a survey of existing document types was performed in order to collectively define the materials that should migrate to any future iteration of Scholarworks. The survey demonstrated overall support for restricting documents within the new IR to those of faculty and student scholarly output (See Background). As a result, many types of

documents such as those relating to University Archives and campus history will be temporarily filtered out of the ScholarWorks migration and remain in DSpace until migrated into the new DAMS and merged with other system-wide digital collections.

Third, digital collections typically include archival and other materials both associated with, and external to, the campus and its history. Therefore, metadata requirements for these materials will also differ from IR materials. This speaks to the need to formalize a metadata working group that accounts for these differences. See Recommendation 3.

The governance body should ensure through equitable representation that all needs of the system are developed such that priorities of IRs and digital collections are equally considered. This recommendation therefore proposes that all interests are represented through special interest advisory groups that make up the governance body. It does not, however, recommend precisely what such a structure should look like. It is known that governance discussions are already in the works and this recommendation is intended to help guide these developments where applicable.

3. **Formalize a *working group* of metadata personnel drawn from participating campuses to lead the continued development of shared metadata practices, formatting, data models and assessments.** This recommendation supports and extends the valuable work already completed by Chancellor's Office (CO) staff and members of the Metadata/Linked Data Interest Group. It also recognizes that, as a fundamental component of digital asset management, metadata applications and iterative assessment are critical to establishing both quality and consistency in our quest to make accessible the objects contained within the DAMS, as well as serve the needs of *all* its users.

   As suggested in recommendation number 2, metadata and its development plays some kind of role (advisory or otherwise) in the overall governance of the system based on the multiple interests it serves between IRs and digital collections. It is the intent of this recommendation that decisions regarding the application of metadata be equitably distributed between participating members, insofar as they are able to contribute, whether as a core member, or designated liaison. To date, this has been covered by the above-mentioned interest group. However, a more formal working group (or similar) may be required to execute future development across the entire consortium. It is envisioned that the most critical work is now through the next year as the group focuses on normalizing structures and workflows. However, future work by the group could come in the form of annual or biennial assessments of campus metadata applications, workflow adjustments, remediation training, annual advisory to the governance body, and more.

4. **Assess the needs, feasibility, and costs associated with adding additional labor necessary to assist CO staff in the management and development of a shared DAMS for digital collections and IRs**. The success of a shared repository will depend

in part on increased commitments of labor and/or expertise potentially through the addition of support staff or third party services.

Funding could be supplied, in whole or in part, by campus savings on existing commitments to third party repository services. Additional funding could be secured by campuses that currently do not have a digital collections repository independent of the existing instantiation of DSpace, but are looking to further develop these services for its users. A group such as a task force charged with surveying and identifying needs that require additional costs associated with a shared DAMS could well serve its development and overall management moving forward.

5. **Assess the willingness and capacity of campuses to help design and program features, within (or external to) the DAMS API that will improve management of the system and user experience**. By definition, open source technologies excel when users make and share their improvements to the platform or application. Therefore, applicable campus personnel should be identified and encouraged to make such contributions when they can, and in the manner of their choosing. However, participation should be voluntary.

   A task force or other small group assembled from existing interest groups could survey campuses for available talent in the areas of programming and user experience design.

APPENDIX A

Glossary

Checksum

When a checksum function is applied to a file the result is a fixed, smaller string of data representative of the whole file. If bits within a file are corrupted or changed, a checksum function can be run again and the resulting string will differ from the original output, indicating a change to the integrity of the file. Checksums are used in digital preservation to determine if a file has become corrupt, usually when uploaded or downloaded to or from a server environment.

Data Integrity

Maintaining the accuracy and longevity of data throughout the digital preservation lifecycle. This is done through various digital preservation strategies including validation audits, fixity checks, inventories and logs, etc.

Digital Asset Management System (DAMs)

A system (usually in the form of a platform or software stack) for storing and maintaining digital files and objects, usually involving a standard set of management services such as organization procedures, storage, retrieval methods, and versioning.

Dublin Core

A metadata schema used to describe digital objects/resources, developed by the Dublin Core Metadata Initiative (DCMI). Dublin Core is less complex than many metadata standards, making it extremely flexible. Many of the metadata fields are repeatable and the overall schema is meant to be syntax independent.

EAD

EAD stands for Encoded Archival Description and an XML standard meant to encode and define the encoding structure of archival finding aids. By encoding a finding aid in XML the expression of the aid is then able to be parsed in a machine readable format.

Fixity

The assurance that a bitstream of a file has remained unchanged, usually using a set of digital preservation tools such as checksums and message digest algorithms.

Institutional Archives

Administrative records whose provenance stem directly from an institutional body as opposed to a single person (personal papers) or other archival collecting areas. Can contain various formats but the items usually pertain to the overall operation and history of a single institution.

JSON

JSON stands for JavaScript Object Notation and is a language agnostic data format (as well as human readable). By transposing a data object into a JSON format it allows for greater transmission of the object between systems, as the format can be parsed by most programming languages.

METS

METS is an acronym for Metadata and Encoding Transmission Standard. METS is a schema that is expressed using XML and is primarily used to describe digital objects. METS documents include metadata concerning the technical and relational aspects of digital objects within a collection or library, ensuring that their persistence is trackable and defined. More information about METS can be found at: https://www.loc.gov/standards/mets/METSOverview.v2.html

MODS

MODS is an acronym for Metadata Object Description Schema. MODS is a schema that is expressed in XML and is primarily used to package bibliographic metadata in a manner that is extensible. Often, MARC 21 metadata is encoded using METS so that metadata is transferable to other library applications. More information about MODS can be found at: http://www.loc.gov/standards/mods/mods-overview.html

OCR

OCR is an acronym for Optical Character Recognition. Using scanning and optical technologies, OCR software takes the text within a document and converts it into a format that can be edited and searched. The text can be digital or handwritten, depending on the software.

OAI-PMH

OAI-PMH is an acronym for the Open Archives Initiative Protocol for Metadata Harvesting. This protocol was created as a mechanism for guiding institutions with online repositories. The protocol offers standards that allow for greater interoperability between systems and system data. For more information see: https://www.openarchives.org/organization/

PREMIS

PREMIS is an acronym for Preservation Metadata Implementation Strategies and encompasses a set of tools used to describe the preservation properties of digital objects. The strategies include a data dictionary, XML schema, and the international standards set forth by metadata experts. For more information see: https://www.loc.gov/standards/premis/

RDF/XML

RDF/XML stands for Resource Description Framework in eXtensible Markup Language. RDF is a data model or framework used to express information about resources in ways that can be interpreted by applications. Where RDF is the framework, XML is a syntax in which the framework can be structured, ultimately allowing for applications to interpret the resource descriptions.

Rights Management

The act of determining the copyright status of a work and describing that status within the metadata record connected to said work when referenced in relation to digital asset management systems.

VRA Core

A metadata standard of the Visual Resources Association to describe works of visual culture. The full schema is available at: https://www.loc.gov/standards/vracore/

WYSIWYG

This refers to “What You See Is What You Get” code editors. When writing in a programming language, WYSIWYG editors display the output of that code to closely resemble what will likely be the final product as opposed to only the coding language in its original form.

APPENDIX B

DIGITAL COLLECTIONS & ARCHIVES SURVEY REPORT

REPORT OF THE

SCHOLARWORKS SWAT TEAM

JANUARY 2019

CALIFORNIA STATE UNIVERSITY

Table of Contents

Background

Of our 23 responses, 20 campuses reported having an digital collections and archives service, with 8 hosted by the Chancellor's Office on DSpace, and 12 hosted by other services. Three campuses do not currently have a digital collections and archives service.

What are campuses spending now on staffing, systems?

Based upon our data, staffing costs for digital collections and archives platforms account for approximately 1.4 million dollars annually ($1,446,592) across the CSU. While the survey provided instructions to account for how to determine staffing costs, the survey did not ask for detailed reporting of campus program costs, such as digitization, technical development, or training.

Our survey additionally asked what campuses were spending on systems (software/hosted platform) for digital collections and archives. Total systems costs accounted for roughly $230,000 ($232,817) annually. Of those costs, campuses with external digital archives systems pay approximately $168,678 annually for services. Respondents who use DSpace for digital archives and collections provided by the Chancellor's Office, and campuses which currently do not have a digital collections and archives service, reported an annual cost of $64,140; however, given that the Chancellor's Office hosted service is currently provided to campuses at no cost, these responses may be related to additional systems used in combination with the Chancellor's Office hosted service, or other systems costs (storage in AWS, etc).

User Stories

To address the questions that we had been tasked with in our committee charge, we first wanted respondents to consider prior system requirements for the Samvera-based service currently in development at the Chancellor's Office. A 2016 System Requirements survey (https://calstate.atlassian.net/wiki/spaces/SCHOL/pages/68517899/ScholarWorks+System+Requirements+Survey) provided list of requirements, which were then reformatted to present user

story statements to the respondents. Respondents were also provided the option to include additional requirements that had not been listed. After examining the results, the statistical differences are minimal. Top priorities include: support for keyword and full-text record searching, support for uploading multiple formats and multimedia files, support for streaming audio and media, and the ability to upload files and records in bulk. At the other end of the spectrum, create lists of published & unpublished works by faculty, provide users the ability to submit their own works, insert graphic widgets such as carousels to promote new additions, initiatives, events, allow faculty, staff, or students the ability to curate their own collections for teaching, and the ability to send email reports as a part of workflows were at the bottom of the rankings.

Top priorities:

	Weighted average	Average	percent very important
[offer keyword search in a search interface]	4.92173913	4.956521739	95.65217391
[provide full text record searching]	4.843478261	4.913043478	91.30434783
[have support for uploading multiple formats and multimedia files]	4.556521739	4.695652174	82.60869565
[enable searching by title, author, and key words for public users]	4.530434783	4.739130435	73.91304348
[provide usage statistics and other reporting on submissions, as through Google Analytics]	4.469565217	4.695652174	73.91304348
[have built in support for streaming capabilities for audio and video]	4.417391304	4.608695652	78.26086957
[offer various filter or faceting capabilities (i.e., by community, by author, by program, by date issued)]	4.269565217	4.565217391	69.56521739
[have built in preservation tools]	4.252173913	4.565217391	65.2173913
[provide an improved, responsive User Interface (UI)]	4.191304348	4.52173913	65.2173913

The top responses identify both discovery and resource-specific functionality as most important. Campuses demonstrate a preference for keyword searching, faceting, and full text searching as most important discovery features. The expressed importance for support for multiple formats, built in AV streaming, and preservation tools are likely reflective of domain-specific requirements for digital collections and archives.

The least important aspects identified relate to management activities. Providing emailed reports, user self-submission and lists of published works for faculty were not as important for digital collection and archives respondents. Stories with the least priority tended to be specific, feature-oriented statements, which may reflect a trend among respondents to prefer comprehensive capabilities (searching full text) over specific feature replication or enhancements (item-based structure to simplify item mapping).

What would we gain by implementing a digital collections and archives platform?

One of the specific questions of the charge, we asked respondents to select which user stories would provide opportunities for the CSU as a whole in digital collections and archives, when compared to how the digital collections and archives services are currently implemented.

When asked “Which do you believe would be most improved by a single digital collections and archives platform?”, the top answers were:

1. Search all communities (18)
2. Built in preservation tools (13)
3. improve discovery of digital objects with internal and external system integrations, such as through OAI-PMH, Calisphere, or DPLA (12)
4. provide an improved, responsive User Interface (UI) (11)
5. Several responses had the same ranking (10) for fifth, including: have built in support for streaming capabilities for audio and video, have modern image display capabilities, including in browser (embedded) zooming and scrolling, provide full text record searching, retain URIs, for instance DSpace Handle identifiers, to offer persistent access to digital objects, provide usage statistics and other reporting on submissions, as through Google Analytics Please see the data.

Additionally, several campuses added potential gains within the survey, with comments including:
“In general, a centralized, single platform would make it easier to develop enhanced tools and services by capitalizing on the expertise and centralizing the work and then pushing it out to all campuses” and, “free up time to develop other features. We could also utilize and share expertise across the system.”

As with our user stories, the most significant improvements identified by digital collection and archives service respondents would be with discovery and domain-related features. A

centralized service would improve community discovery both within the system and with external data services (1,3,5). Additionally, respondents targeted preservation, UI, AV streaming, in-browser image zooming, and other technical features that would benefit from a single service environment. Significantly, it is not clear whether these technical features are considered improvements upon existing systems or services that would need to be replicated. Regardless, respondents likely identified the benefit of centralized technical features with ease-of-implementation in mind, or to reiterate one respondent, " a centralized, single platform would make it easier to develop enhanced tools and services."

What would we lose by implementing a single digital collections and archives platform?

A specific question of the charge, we asked respondents to select which user stories would provide loss for the CSU as a whole in digital collections and archives systems, when compared to how these systems are currently implemented.

When asked "Which do you believe would be least improved by a single digital collections and archives platform?", the top responses were:

1. have the ability to do local administration of site
2. have greater flexibility, customization over the interface (record displays, collection view pages)
3. create custom workflows for managing submissions
4. establish granular permissions for communities and collections
5. configure custom collection pages, exhibit pages, selected works for the campus repository

When considering potential losses for the CSU as a whole in moving to a centralized digital collections and archives environment, loss of local customization was considered most significant (1,2,3,4,5). One important aspect was customization of workflows (3) and granular permissions (4), both which reflected concerns with administrative tools and access restrictions, respectively. These may be identified as losses in comparison to local policies, such as a

campus workflow process within an existing system, or may be considered a loss as relates to existing system capabilities, such as DSpace Authorizations.

Please describe what you believe would be beneficial for your campus by implementing a single institutional repository.

Three threads emerged from the campuses responses regarding the benefits they saw in having a single repository for digital collections and archives. The most consistent response revolved around the idea that a single, shared CSU repository would increase access and visibility, through a variety of opportunities such as cross-campus exhibits, connections and cross-references between similar collections held at different campuses (one campus pointed out the different labor archives at San Francisco, Northridge, and Sacramento), and search results with items from across the CSU. Another theme was an anticipation for better support within a Chancellor's Office-supported environment. Other lesser commonalities among the responses included workflow/management benefits, cost savings, and increased collaboration and sharing among the community of archives and special collections. And the third theme was the improvement of features (or in some cases, have a digital collections system at all), including improved user interfaces, an IIIF viewer, preservation tools, and organizing objects by collection.

What feature(s) of your current platform do you find essential and would need to be retained in a single system?

More data was collected on this last theme when campuses who currently have a digital collections and archives system responded to the question of what features of their current platform would be essential to retain; responses ranged from robust search capabilities, display functionality for different types of objects (books, images, sound, video, etc.), and preservation. Two campuses also pointed out that any feature marked "Very Important" in the user stories section of the survey should be considered essential.

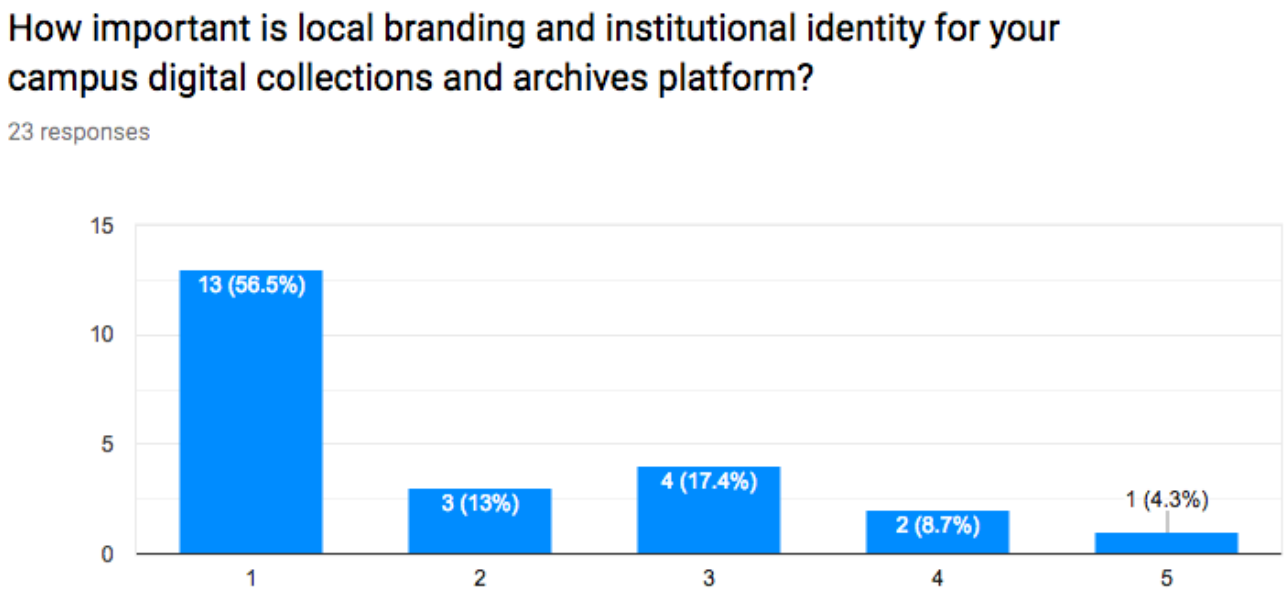
Please describe what you believe would be detrimental for your campus by implementing a single institutional repository.

When prompted to tell what each campus believed would be detrimental with a single digital collections and archives platform, many campuses expressed concern about the loss of administration, customization, and/or control of the repository. Such concerns went beyond branding and interface styling (though concerns for those existed as well)-- campuses identified autonomy over templates, metadata, collection management, and administration of system features as being things that would be detrimental to lose. Another system-related issue revolved around potential confusion over where items are actually physically located, especially in federated search results. But other detrimental possibilities revolved around repository management; campuses expressed concern over how the different campuses and their different needs and priorities might affect how the system is developed and supported. Bottlenecks, striking a balance, reaching consensus, delays in development and varied priorities were all phrases used by the campuses. Similarly, several campuses also worried that a single system would present difficulties in adapting broad solutions to local needs, for example with respect to policies, workflows, collection structures, and metadata and controlled vocabularies. Finally, a few campuses stated that rights management (and how it affects access and re-use policies and workflows) could be detrimental in a single system; though only a handful campuses said this topic could be detrimental, it is worth noting because digital collections and archives can pose different challenges in this area than scholarly works. The intellectual property status of many digital collections is unclear, while scholarly works typically have clear copyright owners.

Importance of Local Branding and Institutional Identity

Maintaining local branding and institutional identity is very important for the campuses (1 being Very Important and 5 being Not at all Important):

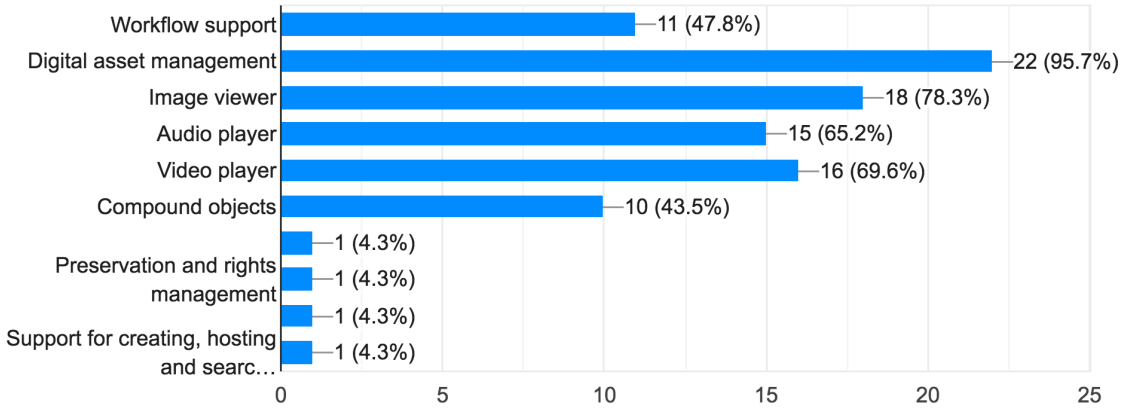
Digital Asset Management appears to be the feature of the highest priority for the Digital Collections and Archives platform, with almost all campuses (22/23) deeming this feature important. Offering a built in image viewer, audio and video players were also deemed important with more than half the campuses expressing interest in these features. Workflow support was also deemed as a feature of high priority with 11 campuses expressing interest in it. Furthermore, having support for compound objects was a feature of interest expressed by a good number of campuses (10).



Features of Highest Priority for a Digital Collections and Archives Platform

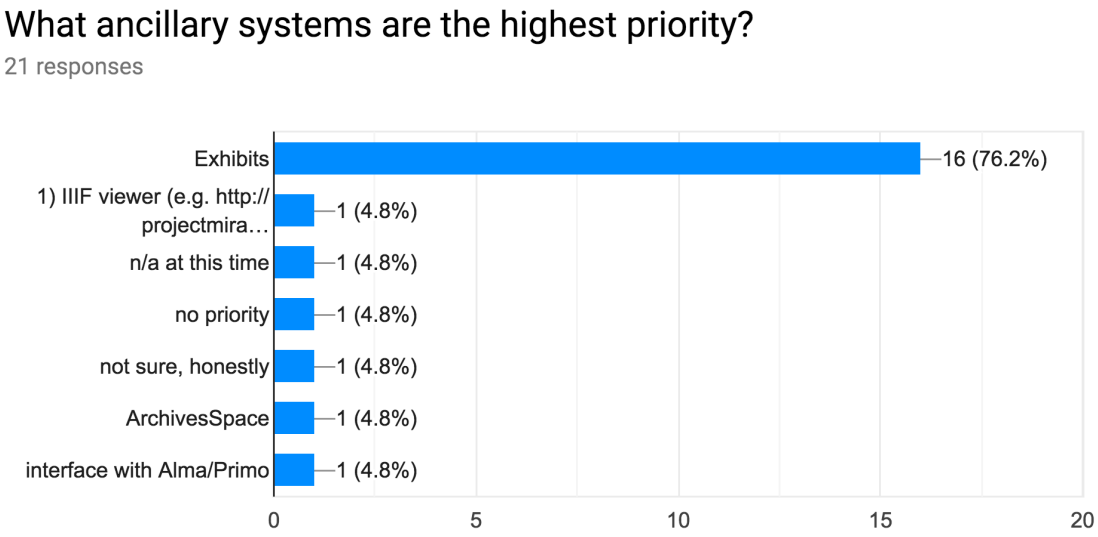
What features are the highest priority for a digital collections and archives platform? (Select all that apply)

23 responses



Ancillary Systems of Highest Priority

Looking at ancillary systems for digital collections and archives is a lot less focused than with the IR survey. The primary response was that exhibits is the most important system. But three respondents said that they didn't know of any or responded with "N/A"



What item types are collected in your digital collections and archives platform?

The top five responses for what types of items are currently collected are:

1. Photographs (22 responses)

Multi-page documents (letters, reports, newsletters)

3. Sound recordings (20 responses)

Videos

Newspapers

Transcripts/interviews

With so many campuses developing diverse digital collections, a single system would likely need to provide support for each of these top 5 types. Other types, such as posters, maps, and architectural drawings are being managed by many campuses, and may need additional functionality and features to support them.

What features/functions to improve on Next Generation Digital Collections & Archives system

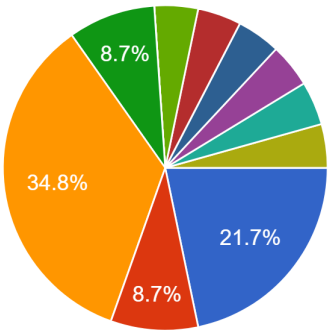
When looking at additional features for a system-wide digital collections and archives platform, there isn't a lot of commonality. Digital preservation is mentioned a few times, as is the ability to bulk ingest items and metadata.

Are there any additional features or functions that you would you like to see in a single system-wide digital collections and archives platform?

11 responses

An easier search and sort function; preservation function
Preservation
OCR; technical metadata for objects
1) User tagging, 2) User description
III standards; III viewer w/ zoom capability
no additional features or functions other than those evaluated in the above sections.
Digital preservation in terms of stability, format migrations, fixity checks/checksum, etc.
Ability to bulk upload and edit globally. Consistent metadata. Templates for various item types.
Abillity to ingest and preserve born digital materials.
Batch upload
I think this survey has them well-covered.

What platform(s) does your library use to manage and provide access to digital collections and archives?

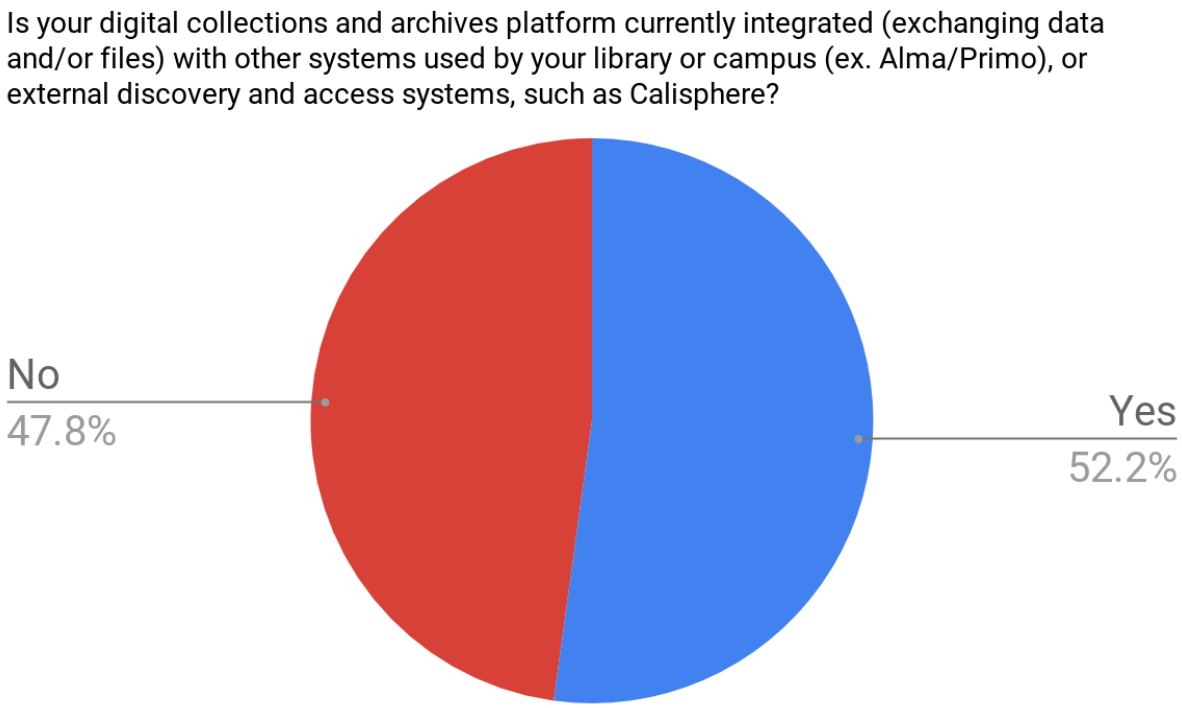


The top 2 current platforms being used are CONTENTdm and DSpace (likely ScholarWorks). Much lower on the list, with 2 responses each, are Islandora and DigitalCommons. With 1 response each are iBase, Omeka, SF State's locally developed DIVA platform,

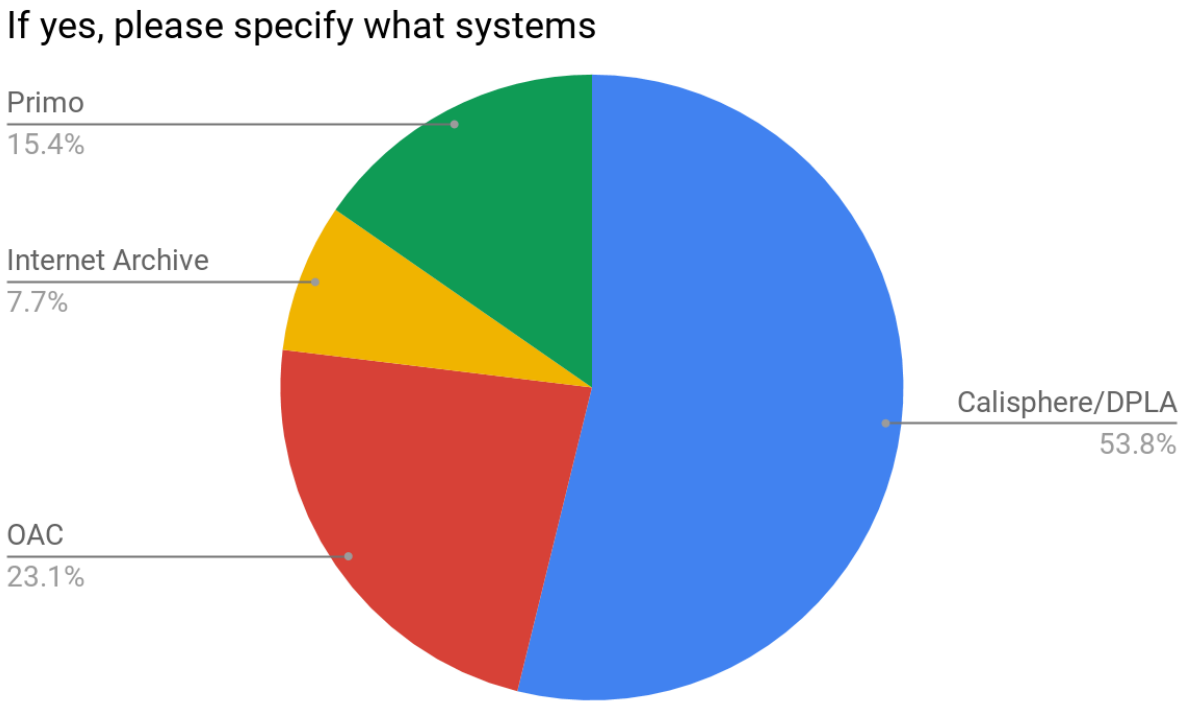
ContentPro IRX (iii), and Microsoft Access. A single system would face significant challenges normalizing, standardizing, and preparing content and metadata to be migrated from a such a diverse set of systems. Each system may also have its own set of features, which in aggregate may be challenging to implement in a single system.

Integration with Other Systems

Integrations chart #1



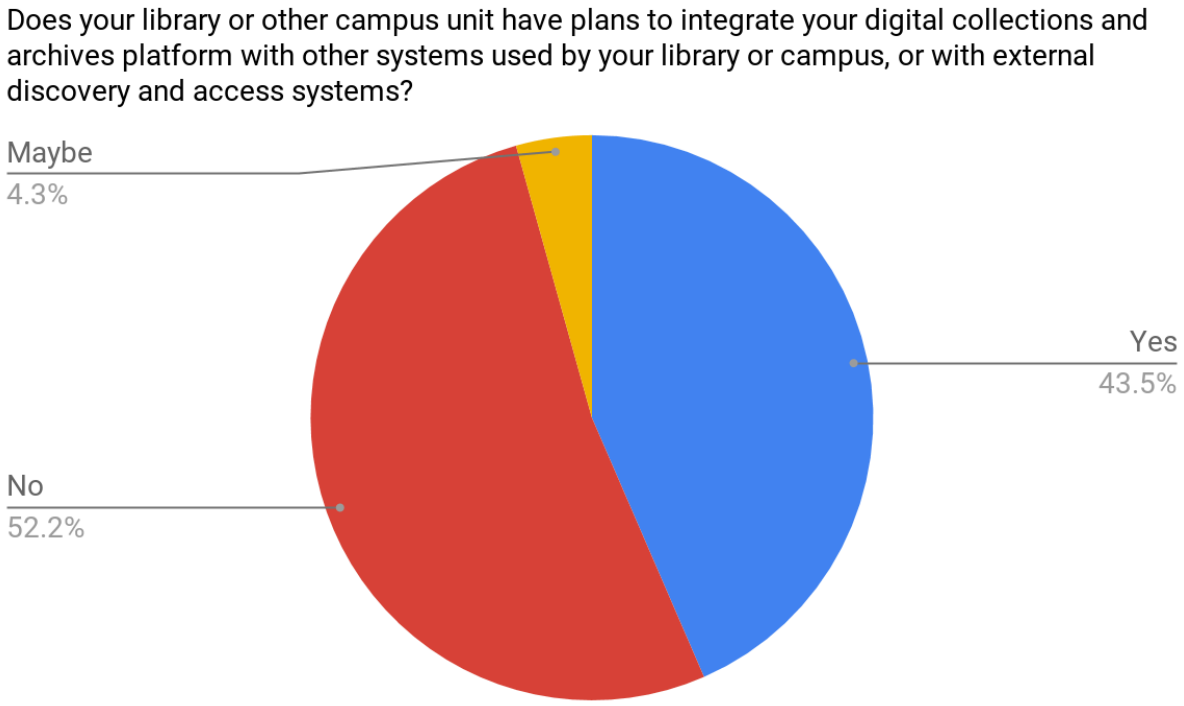
Integrations chart #2



*The above chart represents 13 responses

**OAC = Online Archive of California // DPLA = Digital Public Library of America

Integrations chart #3



The YES responses in Integrations Chart #3 roughly correspond to the same respondents who answered NO in Integrations Chart #1. However, there are a few campuses that do not have current integrations and who also have no integration plans; and at the same time, there are 1 or 2 campuses that have current integrations, and want more! Many of the NO responses in Integrations Chart #3 are satisfied with their current Calisphere/DPLA integration, and have no plans to expand further.

In any case, it is clear that there is significant (if not wildly diverse) integration activity and plans, even for campuses that do not currently have a digital collections and archives system. A single CSU system would need to take these integrations into account.

Additional Challenges and General Comments

A good number of campuses have deemed preservation as an important feature of a Digital Collections and Archives Platform. The general consensus is that CSU campuses would benefit from a centrally hosted platform when considering the need for offering and implementing digital

preservation. Other advantages of such centrally hosted and supported platform include better handling of A/V materials, simplifying ingest and offering increased control over digital objects. On the other hand, being able to retain local control of branding features, customizing metadata schemas, managing workflows, and being able to configure embargoes are some of the expressed concerns with moving to a single, centrally hosted system.

APPENDIX C

External Environmental Scans and Reports

Stein, Ayla, & Thompson, Santi. (2015). *Taking Control: Identifying Motivations for Migrating Library Digital Asset Management Systems*. D-Lib Magazine, 21(9). Retrieved from http://www.dlib.org/dlib/september15/stein/09stein.html

UC Libraries DAMS Technology Team. (2016). *UC Libraries DAMS Technology Report: Assessment of a Long-Term Solution for the UC Libraries Systemwide DAMS*. Retrieved from https://wiki.library.ucsf.edu/display/UCLDTP/UC+Libraries+DAMS+Technology+Project+Home?preview=/380146695/397772708/DOC_DAMS_FinalReport_01Nov2016.pdf#UCLibrariesDAMSTechnologyProjectHome-FinalReport

Broadwell, P., & Childress, D. (2018). *Digital Library Systems and UCLA: An Environmental Scan with Suggestions for Near- and Medium-Term Technology Strategies and User Stories / Functional Requirements*. Los Angeles, CA: University of California Los Angeles.*

UC Libraries Linked Data Project Team. (2018). *UC Linked Data Project Team Report: Systemwide Work, Opportunities, and Recommendations.* Retrieved from https://libraries.universityofcalifornia.edu/groups/files/doc/docs/UC%20Linked%20Data%20Final%20Report_01Oct2018.pdf

*Report not available to the public