

SWAT Team Report

California State University

February 19, 2019

- Alyssa Loera, Pomona
- Andrew Weiss, Northridge
- Carmen Mitchell, San Marcos
- Curt Asher, Bakersfield
- Dana Ospina, Dominguez Hills
- David Walker, Chancellor's Office
- Karen Schneider, Sonoma
- Kevin Cloud, Chancellor's Office
- Renaldo Gjoshe, Fresno
- Tracy Elliott, San Jose
- Zach Vowell, San Luis Obispo

EXECUTIVE SUMMARY **4**

Recommendations 6

REPORT 1: INSTITUTIONAL REPOSITORY SURVEY **8**

Background 8

What are campuses spending now on staffing, systems? 8

User stories 8

What would we gain by implementing a single institutional repository? 10

What would we lose by implementing a single institutional repository? 11

Please describe what you believe would be beneficial for your campus by implementing a single institutional repository. 12

Please describe what you believe would be detrimental for your campus by implementing a single institutional repository. 12

How important is local branding and institutional identity to campuses? 14

What features are the highest priority for an institutional repository? 14

What ancillary systems are the highest priority? 15

What item types are collected in your institutional repository? 16

Integration with other systems 17

Additional comments & challenges 17

REPORT 2: DIGITAL COLLECTIONS & ARCHIVES SURVEY **18**

Background 18

What are campuses spending now on staffing, systems? 18

User stories 19

Top priorities 19

What would we gain by implementing a digital collections and archives platform? 20

What would we lose by implementing a single digital collections and archives platform? 21

Please describe what you believe would be beneficial for your campus by implementing a single institutional repository. 22

What feature(s) of your current platform do you find essential and would need to be retained in a single system?	23
Please describe what you believe would be detrimental for your campus by implementing a single institutional repository.	23
Importance of local branding and institutional identity	24
Features of highest priority for a digital collections and archives platform	25
Ancillary systems of highest priority	25
What item types are collected in your digital collections and archives platform?	26
What features/functions to improve on next generation digital collections & archives system	27
What platform(s) does your library use to manage and provide access to digital collections and archives?	28
Integration with other systems	29
Integrations chart #2	30
Integrations chart #3	31
Additional challenges and general comments	32
REPORT 3: CONSORTIA ENVIRONMENTAL SCAN	33
Background	33
What other consortia have a centrally-hosted IR or digital library?	33
What kind of staffing do they have for their IR or digital library?	34
Are the positions centrally funded, or do campuses contribute?	35
Conclusions	46
APPENDIX A: Questions posed to participating consortia	48
APPENDIX B: CSU Institutional Repository Collections Totals	49
REPORT 4: TECHNICAL CONSIDERATIONS FOR A SHARED INSTITUTIONAL REPOSITORY	50
The need for multitenancy	50
Demo system	50

Executive Summary

The SWAT Team was charged by COLD with reexamining the requirements and resources needed to support the ScholarWorks project. In particular, the team examined the possibility of a single, system-wide institutional repository (IR) rather than separate, campus-facing institutional repositories, as we have today.

To answer that question, the SWAT team completed several tasks:

- Two internal surveys of the CSU libraries, gathering feedback on the perceived impact of a single-instance IR, as well as a broader needs assessment for both IR and digital archives projects (**see Reports 1 & 2 below**).
- Interviewed nine consortia with a shared institutional repository or digital library to understand their current organization (**see Report 3 below**).
- Developed a [demo system](#) to explore the technical feasibility of a single-instance repository (**See Report 4 below**).

The pros of the single-instance model include:

- Fewer server resources, reducing operating costs
- Simpler configuration
- Concentration of limited CO programming and design resources on a single system, both for the initial implementation and future enhancements and upgrades
- Potentially greater visibility of content

The cons include:

- Limited institutional branding, which was deemed of high importance by most campuses
- No local customizations to metadata schemas, aside from controlled vocabularies (e.g., for college and department names)
- The need to merge / re-migrate data into a single system

- The likely need to relocate material currently in ScholarWorks that is deemed out-of-scope to another system.

That last point is not strictly necessary, but, taken together with the desire of some campuses to replace CONTENTdm with ScholarWorks, or even to start a new digital archives from scratch using ScholarWorks, it points to the need for a separate instance of Hyrax to support a centrally-hosted digital archives and collections service. The work needed to implement such a service is probably equal to that of the IR, as digital archives have unique requirements.

Finally, outside of the core institutional repository and digital archives systems, the campuses identified three ancillary services as being of highest importance:

- Faculty profiles
- Open access publishing
- Exhibits

The Chancellor’s Office already [hosts OJS](#) for journal publishing, and there are open source options for monographs, conferences, exhibits and faculty profiles, although the options for faculty profiles are both limited and complex to implement, and the CSU may actually find developing a light-weight, bespoke application based around ORCID the better approach.

While open source systems cost nothing to license, the work needed to implement these various services, migrate data from any legacy systems, and provide ongoing support is significant. As the bulk of this work lies in implementation and upgrades, some tasks could potentially be sped up by additionally hiring short-term contractors. We estimate the following *minimum* staffing may be sufficient:

System	Minimum FTE staff
Samvera -- Institutional repository	1.0
Samvera -- Digital archives	1.0
Open Journal System	.25

Open Monograph & Conference Systems	.25
Faculty profiles	1.0
Exhibits	.5

The CO currently has 1.5 FTE devoted to ScholarWorks and an annual operating budget of about \$30k, covering servers, storage, and long-term preservation costs. We estimate that budget can accommodate upwards of 7 TB to 14 TB of data, depending on how frequently files are downloaded. Current usage in ScholarWorks is about 5 TB, and is poised to grow rapidly. Additional funding will be needed.

Recommendations

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

4. **Fund an additional position at the Chancellor's Office.** An extra position would provide both greater continuity of service during any staffing turn-over at the CO as well as address some of the systems and services outside of the IR listed above, based on what COLD determines are priority needs.

5. **Explore new revenue streams / funding opportunities.** There is clearly a need for additional funding to support ScholarWorks. Report 3 below provides several examples of how other consortia have done this. The proposed governance structure or a task force should examine this issue and provide a concrete recommendation.

Report 1: Institutional Repository Survey

Background

Of our 24 responses, most campuses have an institutional repository service, with 16 hosted by the Chancellor's Office on DSpace, and seven hosted by other services. One campus reported not currently having an institutional repository.

What are campuses spending now on staffing, systems?

Staffing costs for CSU-wide institutional repositories account for approximately 1.2 million dollars annually (\$1,197,721). 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 or training.

Our survey additionally asked what campuses were spending on systems (software/hosted platform) for the institutional repository, including digital preservation costs. Total systems costs accounted for roughly \$220,000 (\$219,959) annually. Of those costs, campuses with external IR systems pay approximately \$181,959 annually for services. Some respondents who use the IR service provided by the Chancellor's Office reported an annual cost of \$38,000; however, given that the Chancellor's Office hosted service is offered at present at no cost, these responses may be related to additional systems used in combination with the Chancellor's Office hosted service.

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

provided a 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: supporting uploading multiple file formats, keyword searching in a search interface, filtering or faceting capabilities, providing usage statistics and full text record searching. At the other end of the spectrum, faculty, staff, or student collection curation, widgets such as carousels to promote new additions, and greater flexibility, customization over the interface were at the bottom of the rankings.

	Weighted Average	Average	Percent of responses "Very Important"
[have support for uploading multiple file formats and multimedia files]	4.641666667	4.791666667	83.33333333
[offer keyword search in a search interface]	4.566666667	4.75	79.16666667
[offer various filter or faceting capabilities (i.e., by community, by author, by program, by date issued)]	4.491666667	4.708333333	75
[provide usage statistics and other reporting on submissions, as through Google Analytics]	4.416666667	4.666666667	70.83333333
[provide full text record searching]	4.3	4.583333333	66.66666667
[integrate with common cross-campus authentication methods (CAS/Shibboleth/LDAP)]	4.225	4.541666667	66.66666667
[upload files and records in bulk]	4.208333333	4.541666667	62.5
[review submissions to make edits to metadata and files]	4.166666667	4.5	62.5
[enable searching by title, author, and key words for public users]	4.041666667	4.458333333	50

One of the priorities mentioned in the survey, integration with Alma/Primo was a high priority for several campuses. Some of the respondents mentioned that they currently have integration with their discovery system, and would need this as a feature in any new system.

While it is difficult to identify trends, the top responses were user stories related to discovery within the IR. Whether that included searching, filtering, or faceting within an interface, supporting multiple file formats, streaming for AV or even reporting of usage statistics, respondents identified requirements which enabled users, or improved upon users' ability in existing systems, to find digital objects or to deposit a variety of digital object formats, as most important.

Additionally, it is challenging to determine trends for the least important responses. However, one trend for the responses that were considered least important is that they tended to reference specific features rather than broadly defined features (consider, 'offer keyword search in a search interface', the second most important user story, compared to 'insert graphic widgets such as carousels to promote new additions, initiatives, events', the second least

important). This trend is reinforced in later sections of the survey where respondents reported that the “most important concern is getting repository up and running, additional features/systems can come later.”

What would we gain by implementing a single institutional repository?

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 IR systems, when compared to how the IR is currently implemented.

The top five responses for “What would we gain by implementing a single institutional repository?” are:

1. Search across all communities. (18)
2. Improve discovery of digital objects with system integrations, such as Alma/Primo, ORCID or SWORD. (16)
3. Provide a good Application Program Interface (API) for data exchange and integrations. (13)
4. Have built in support for streaming capabilities for audio and video. (12)
5. Offer system-guided accessibility checking/editing. (12)

Several campus respondents provided free text answers as well, including: “Key benefit to having a single instance is focusing on delivering quality features,” and, “provide users with a larger pool of material to search and browse.”

As with our previous user story section on the most important aspects of the IR, the most significant benefits (1,2,3, above) identified within the system relate to discoverability within the IR. By having all CSU collections available within a single repository, respondents suggest a net benefit by providing users single service entryway into the quantity of CSU IR data and

materials. At the same time, respondents seem to suggest that the quality of technical features (4,5) would be improved in a single IR.

What would we lose by implementing a single institutional repository?

A specific question of the charge, we asked respondents to select which user stories would provide loss for the CSU as a whole in IR systems, when compared to how the IR is currently implemented.

The top five responses for “What would we lose by implementing a single institutional repository?” are:

1. Have the ability to do more local administration of site (17)
2. Have greater flexibility, customization over the interface (record displays, collection view pages) (14)
3. Create custom workflows for managing submissions (14)
- 4./5. (A three-way-tie, 11) Have the ability to customize the user interface.
 - Customize and control templates and styling for the repository.
 - Provide authority control to allow maintaining controlled vocabulary locally.

Comprehensively, campus responses identified customization as a net loss when considering a single IR service. Whether taken from the perspective of administrative functionality (1,3) or a UI perspective (2,4/5), campuses suggest that a move to a single service would deter localization options available within current systems.

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

Campuses consistently expressed that technical support for a single repository would be improved. Not only would the Chancellor's Office be able to focus on one installation (with its routine updates and maintenance), but also individual campuses would be able to form "dedicated development teams" and work collectively towards integration, feature development (such as a streaming server, preservation tools, etc.) and collection support. Not only would support improve, but implementation of Samvera would be faster, which several campuses looked forward to eagerly. Another theme that emerged was that a single repository could foster "better coordination across campuses," which could result in improved, shared, and unified documentation, workflows, data models, metadata quality, and standards. Best practices and communities of practice could emerge. Such a "tighter networking of colleagues" could have related benefits, such as demonstrating a more unified front on open access. In addition to these two staffing considerations, campuses also consistently responded that a single repository would improve access, broaden the audience for CSU research, and more easily facilitate demonstration of CSU accomplishments. A smaller (but still significant enough) number of campuses said in general a single repository would make more efficient use of CO resources, concentrating those resources, and spreading the benefits of a robust IR system to the small campuses as much as the large ones.

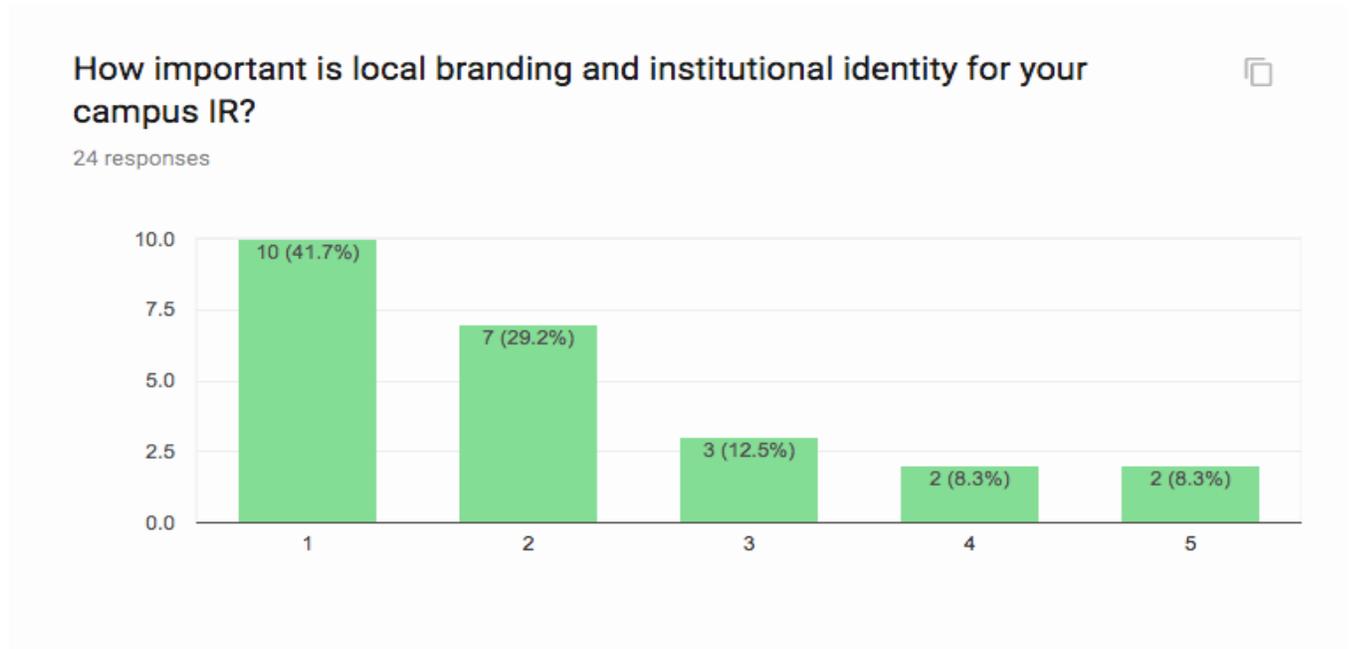
Please describe what you believe would be detrimental for your campus by implementing a single institutional repository.

A general feeling that loss of local control (in whatever area of the repository) was pervasive in the campuses' responses. One campus ventured one word to describe what would be lost in a single repository scenario: "autonomy." Many campuses went further and spoke of the loss of customizations for interfaces and branding as being specifically detrimental. Related to this apprehension towards loss of control were the many responses which discussed potential

difficulties around modifying and adapting local workflows, policies, metadata requirements, software applications, and other existing functionality to a single repository structure. In some cases, these responses had less to do with repository functionality, and more to do with the work it would take to get everyone on the same page (this was especially pointed out for metadata). And granting that such local needs would be entertained by the single repository administrators, several campuses expressed uncertainty to what extent local campus resources would be needed to realized optimal functionality; and given that the resources are present, other campuses said they foresaw detrimental campus disagreements and competing priorities hindering consensus on the "path forward." In addition to these broadly felt concerns, some campuses brought up the potential loss of control over what types of objects are in the IR, others foresaw challenges in reporting functionality and formats that would satisfy all, while others saw little detriment to a single repository, with one campus stating: "As long as we have a working IR, we'll be happy."

How important is local branding and institutional identity to campuses?

Maintaining local branding and institutional identity is very important for the campuses



(1 being Very Important and 5 being Not at all Important):

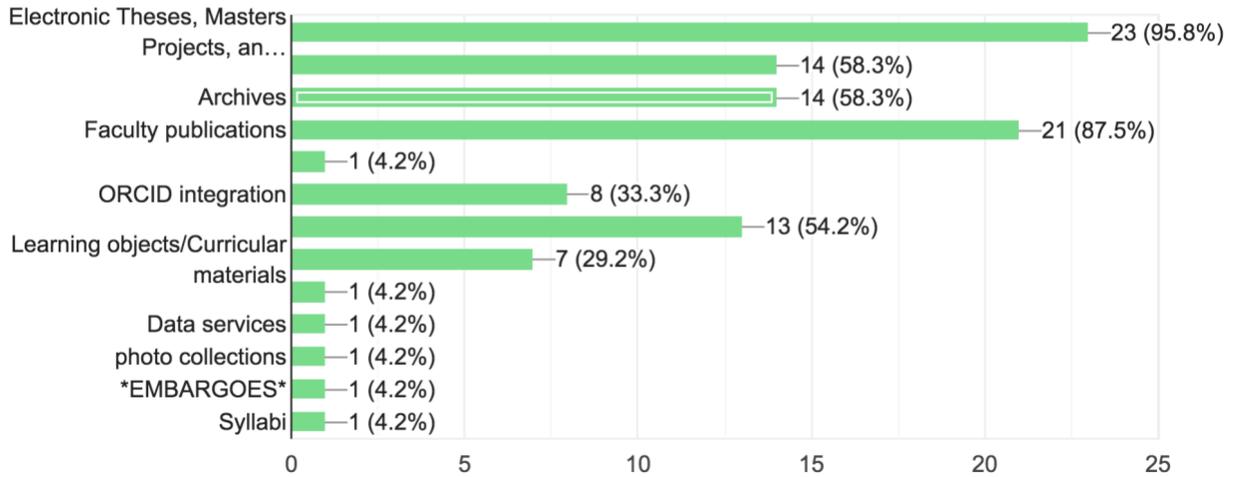
What features are the highest priority for an institutional repository?

Survey results show that hosting Electronic Theses, Masters Projects, and Dissertations (ETDs) is the highest priority for an IR, with almost all campuses (23) selecting this feature. Faculty publications came in at a close second with 21 campuses wanting to see this feature implemented. Other notable features of high priority were Student works (not ETDs) with 14 campuses deeming this feature important; Archives (14 campuses); Learning Objects and Curricular Materials (7 campuses).

Campuses also expressed interest in having the ability to mint DOIs (13 campuses) and being able to have an integration with ORCID (8 campuses).

What features are the highest priority for an institutional repository? (Select all that apply)

24 responses

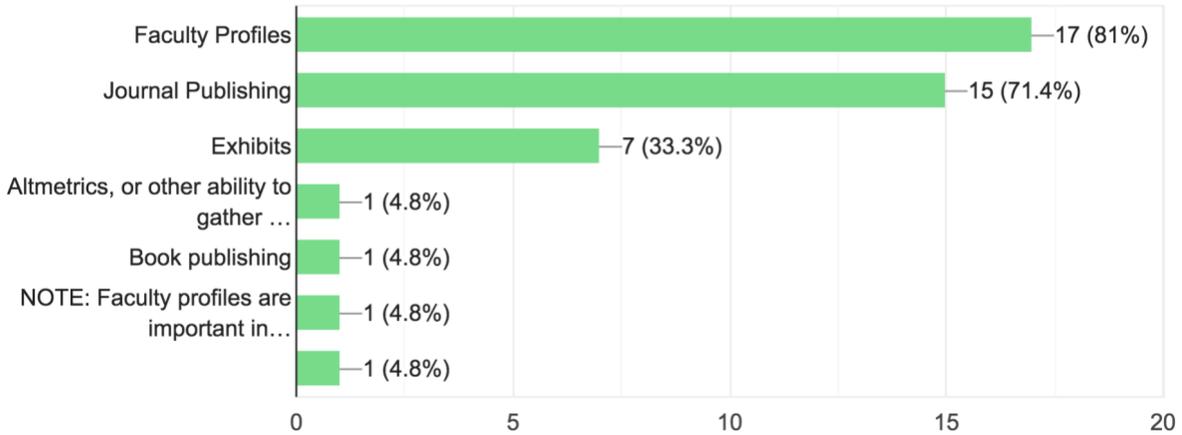


What ancillary systems are the highest priority?

When asked about what ancillary systems are the most important to the CSU IR, responses were pretty consistent. Faculty profiles are the most important priority, with journal publishing and exhibits following at a close second and third.

What ancillary systems are the highest priority?

21 responses



What item types are collected in your institutional repository?

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

1. Electronic Theses, Masters Projects, and Dissertations (ETDs)
2. Faculty publications
 - Audio-visual materials
 - Campus journals, newspapers, newsletters
3. Poster presentations
 - University Archives materials
 - Images

It is no surprise that ETDs and faculty publications are at the top of this list. It is noteworthy that audio-visual materials rank as high as faculty publications, and images are nearly as prevalent (at a 62.5% response rate, as opposed to 79.2% for audio-visual). But perhaps most interesting is the presence of University Archives and campus publications (many of which may

be considered a part of University Archives) also in the same company as the scholarly output traditionally associated with IRs.

Integration with other systems

Every campus reported that Alma/Primo integration was important for their IR. In some cases, Alma/Primo integration is feature or area of potential integration for the future IR; for others, the IR would be expected to reproduce existing services that deposit records into campus discovery systems. In addition, ePortfolio, Interfolio, eFaculty 180, and PeopleSoft were identified by campuses as important potential system 'pipes.' Several campuses expressed a desire to have ORCID integrations available within the new system. Data from the user stories section reiterates the importance of system integration within the IR, with an average response of 'somewhat important.'

Additional comments & challenges

Some commonalities emerged from the additional comments and challenges. A few campuses expressed a desire to see a new repository system "up and running", and rating that need higher than anything else. As one campus stated it: "Getting something up is more important than a fancy feature-rich interface." While a few other campuses want to make sure that current IR features are supported in a new platform. A handful of other campuses made comments about costs related to a single repository scenario (one anticipating higher costs, and another speculating about lower costs). Finally, a couple of campuses made comments regarding the Chancellor's Office's role in supporting the IR and the work that depends on it.

Report 2: Digital Collections & Archives Survey

Background

Of our 23 responses, 20 campuses reported having a 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](#) 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.

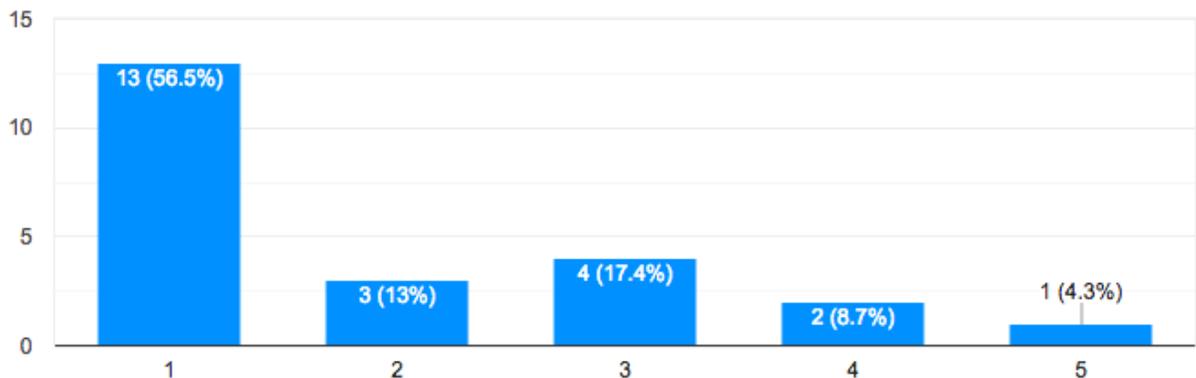
Importance of local branding and institutional identity

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).

How important is local branding and institutional identity for your campus digital collections and archives platform?

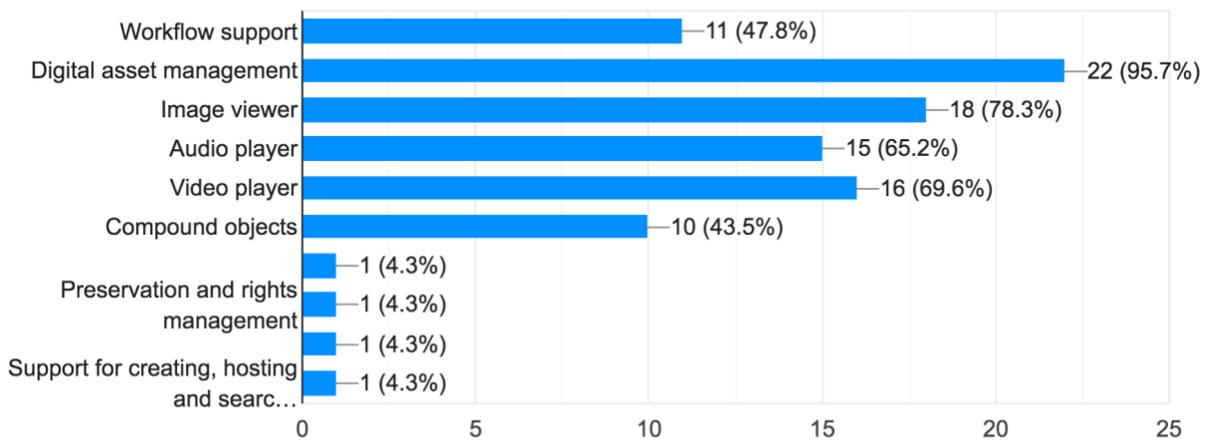
23 responses



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

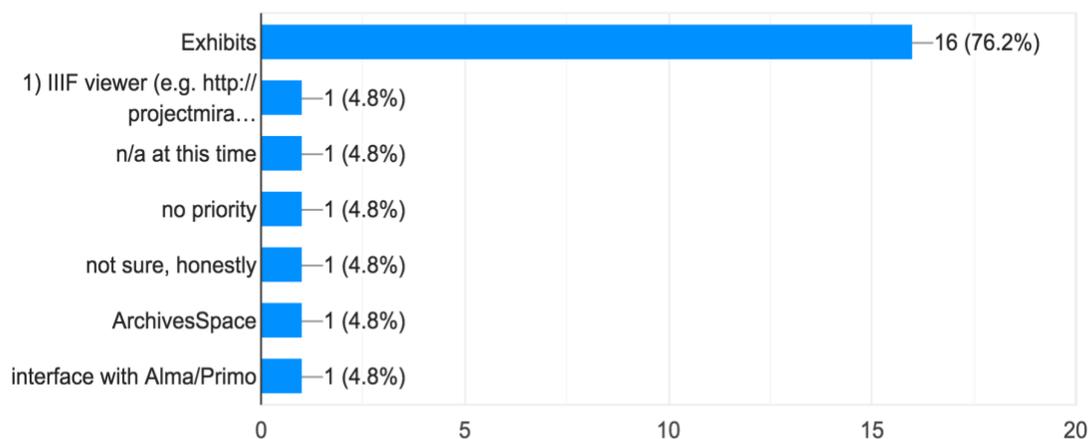


Ancillary systems of highest priority

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 ancillary systems are the highest priority?

21 responses



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)
2. 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.

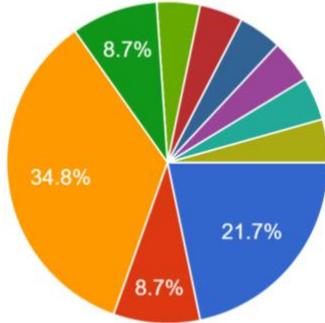
Ability to bulk upload and edit globally. Consistent metadata. Templates for various item types.

Ability 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?



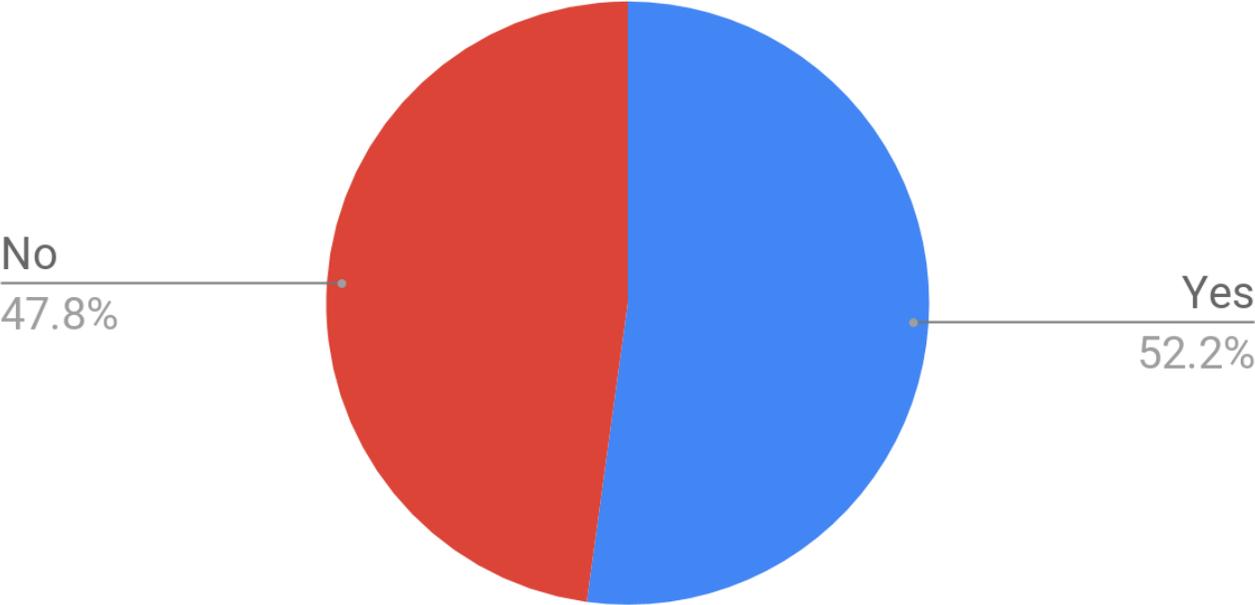
The top two current platforms being used are CONTENTdm and DSpace (likely ScholarWorks). Much lower on the list, with two responses each, are Islandora and DigitalCommons. With one response each are iBase, Omeka, SF State’s locally developed DIVA platform, ContentPro IRX (iii), and Microsoft Access. 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

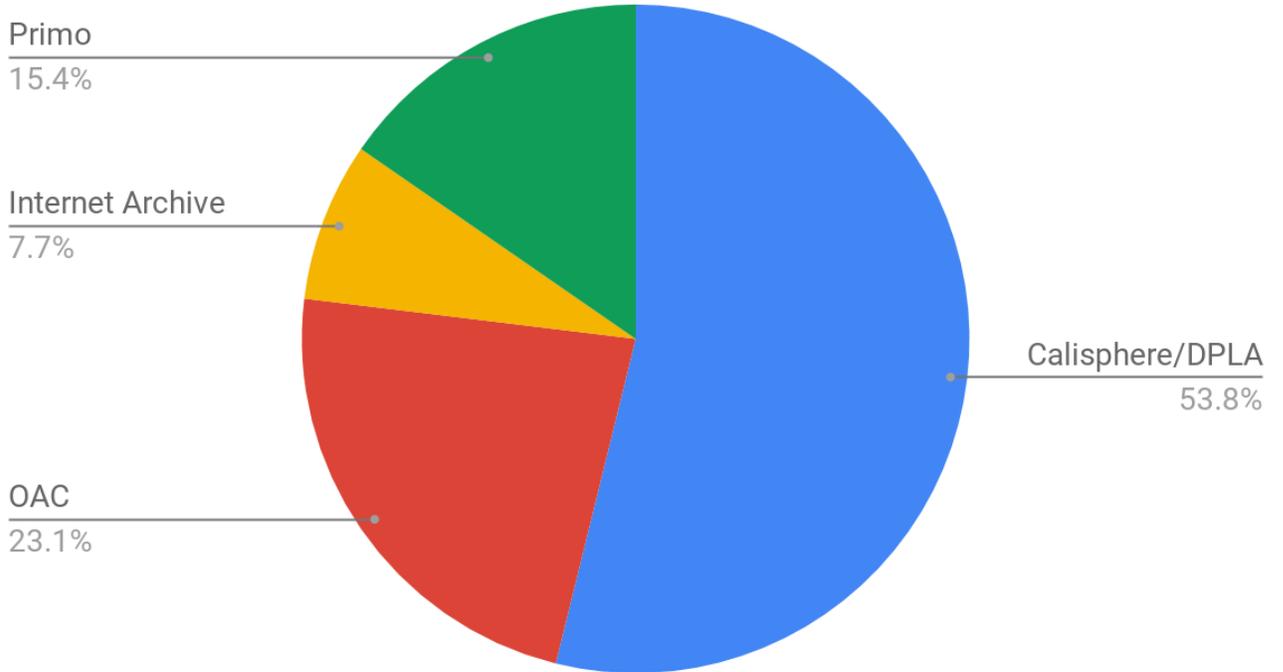
Integrations chart #1

Is your digital collections and archives platform currently integrated (exchanging data and/or files) with other systems used by your library or campus (ex. Alma/Primo), or external discovery and access systems, such as Calisphere?



Integrations chart #2

If yes, please specify what systems

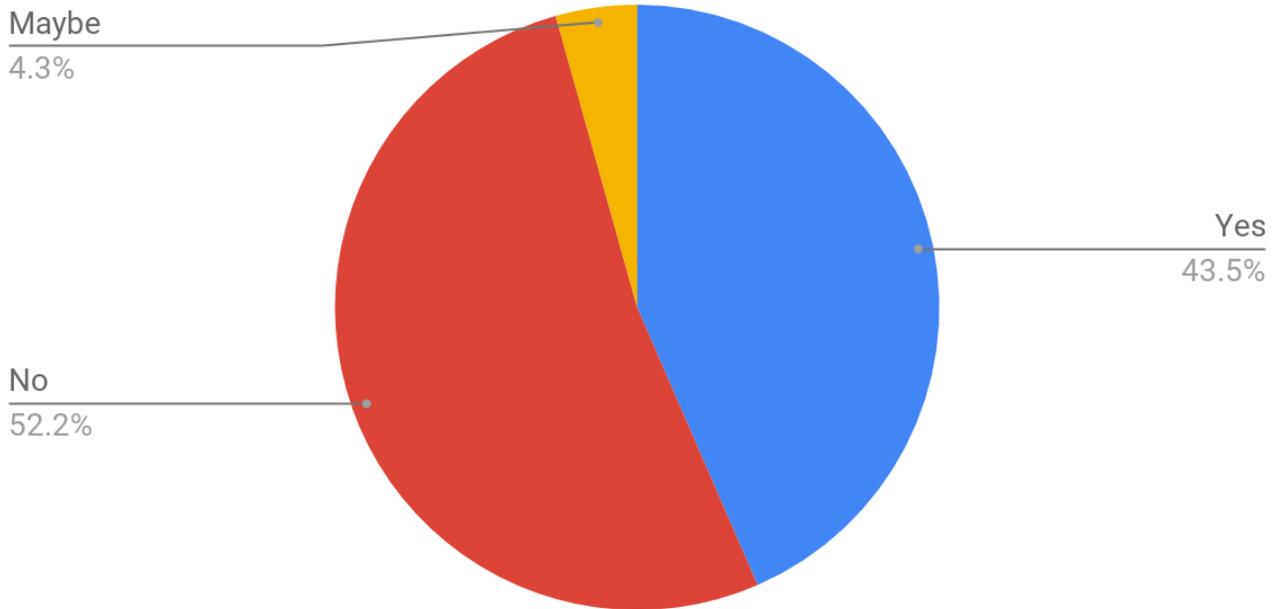


*The above chart represents 13 responses

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

Integrations chart #3

Does your library or other campus unit have plans to integrate your digital collections and archives platform with other systems used by your library or campus, or with external discovery and access systems?



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 one or two 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.

Report 3: Consortia Environmental Scan

Background

The SWAT Team conducted an environmental scan of consortia that currently implement or have, in the recent past, implemented a single, centralized institutional repository and/or digital library. After researching and identifying consortia that implement organizational models ranging from membership to system-based to resource sharing, we either interviewed or corresponded with nine consortia, and asked all to respond to a set of questions (**See Appendix A**) regarding staffing, cost and funding models, platform selection, metadata, success, and satisfaction. Our interactions with the institutions varied: We conducted hour-long interviews with three participants (CDL, CUNY, TDL); these interactions yielded the most information and insight. A couple of the remaining institutions provided us with substantive correspondence (Oregon Digital, University of Missouri), and the remaining sent responses to our questions.

The report that follows responds to the questions posed to us by the original SWAT team charge. Additionally, we include relevant information offered by participants in response to the questions we designed, some of which were not directly addressed in the charge.

What other consortia have a centrally-hosted IR or digital library?

While there are numerous centrally-hosted systems in existence ([OpenDOAR](#) provides information on many), we focused on contacting consortia that we believed would provide the most salient and useful information for the issues under consideration by COLD and the SWAT Team. The nine consortia who responded to our request were:

- [California Digital Library](#)
- [City University of New York](#)
- [Mountain Scholar: Digital Collections of Colorado and Wyoming](#)

- [Oregon Digital](#)
- [Texas Digital Library](#)
- [Washington Digital Libraries Consortium Digital Collections](#)
- [State University of New York](#)
- [OhioLink - Electronic Thesis and Dissertation Center](#)
- [University of Missouri](#)

What kind of staffing do they have for their IR or digital library?

After reviewing the staffing models of the consortia with shared institutional repositories and/or digital libraries, three themes emerged.

1. **Large systems with separate and distinct centralized organizations.** Consortia such as OhioLink, Washington Research Libraries Consortium (WRLC), the California Digital Library, and Texas Digital Library have robust staffing models that serve not only the IR, but several other digital platforms and even in some cases, shared library services platforms. In these cases, the direct support for the IR is a fraction of their work.

IR name	Total FTE for all projects	Total FTE devoted to the IR or DL	IR Platform	Institutions served
WRLC	20	.3 IR only	Islandora	9
Texas Digital Library	9	9	DSpace	16
CDL eScholarship	92	11 IR only	Homegrown	10
OhioLink ETD	17	3	Homegrown	33

2. **University based.** Consortia such as Mountain Scholar (which includes institutions in both Colorado and Wyoming) and Oregon Digital. By hosting the systems locally, the universities are able to utilize existing staff, such as developers and managers who oversee multiple local projects. Colorado State University hosts Mountain Scholar, whereas University of Oregon and Oregon State share staff for Oregon Digital Library. Both of the Oregon institutions host their own IRs, only sharing a digital collections platform. Colorado State's Mountain Scholar platform also serves as a digital library.

IR name	Total FTE for all projects	Total FTE devoted to the IR or DL	IR Platform	Institutions served
Mountain Scholar	8 DL only	5	DSpace	10
Oregon Digital	12	4 DL only	Samvera	2

3. System office-based. Systems such as CUNY Academic Works and SUNY Digital Repository represent multi-campus university systems and have systems offices with varying degrees of centralized support. They are both systems very similar to the CSU.

IR name	Total FTE for all projects	Total FTE devoted to the IR or DL	IR Platform	Institutions served
CUNY	Not available	1	Bepress	25
SUNY	Not available	.25	DSpace	40

NOTE: SUNY's repository is hosted and managed by Atmire, a DSpace contractor.

Are the positions centrally funded, or do campuses contribute?

This section will examine the governance and funding models of several consortial-sized repositories in order to help better describe the cost of sustaining these repositories. We believe it is useful for us all to see how these repositories developed over time, the steps taken to development, the roadblocks they endured and ultimate ongoing and sustainable successes they have accomplished. The platform, and whether it is proprietary or open source, is also provided in this section.

California Digital Library (CDL): *eScholarship*

- Platform: unique homegrown system
- Platform purpose: Calisphere is a sister system for unique digital collections whereas eScholarship acts as a repository for ETDs and faculty scholarship, as well as an open access publishing platform. The IR and the DL are separate platforms.
- Collections scope: Institutional Repository, faculty publications and ETDs, student capstone projects, and conference proceedings

- Collection size: 200,000+

Background and governance / funding

CDL's *eScholarship* includes the **10 campuses** of the UC system. *eScholarship* is but one section of the much larger California Digital Library. By itself, *eScholarship* staffs approximately **12** people. It receives its funding directly from the CDL, which itself is a unit within the department of Academic Personnel and Programs, in the Division of Academic Affairs at the UC Office of the President. Campuses don't appear to pay extra fees for use of the repository.

On a special note: As of December 2017, the CDL has been awarded funding of **\$21,712,755**, some of which has been used over the years to fund specific *eScholarship* projects.

Lessons / takeaways

The funding and governance structure at the CDL is completely different than that of the CSU. Services provided are funded through the UC Office of the President. Grants seem to be a major source for funding development projects, something that the CSU has not been active in doing for the repository. We can mimic, but the overall strong governance structure of the CDL is what makes the *eScholarship* repository so successful and so *sustainable*. We would be well-advised to consider a different model to imitate, since the parameters are so different. The lessons we take from CDL might not translate well to the CSU.

City University of New York (CUNY)

- Platform: Bepress
- Platform purpose: Digital Collections and Institutional Repository
- Collections scope: Hosts open educational resources, ETDs, and archival collections from the various contributors. Though not in an ideal manner (mostly formatted as PDFs). The CUNY documentation doesn't seem to address digitized special collections materials but it is possible that each contributor decides how to use the system. Acts like a catch-all

repository with a lack of customization on the front end (meaning it is less like a digital library).

- Collection size: 19,000

Brief:

CUNY includes **25 campuses and 31 libraries (four-year and CC)** in the metropolitan New York City region; the repository is coordinated by the Office of Library Services at CUNY Central. Works are selected and deposited by the individual campuses in consultation with the Office of Library Services.

Background and governance / funding

Funding is central. It costs approximately **\$250,000** per year for a single-instance Bepress repository. Staffing includes one Scholarly Communication Librarian and 20 campus Academic Works administrators. Bepress provides tech support for the repository, allowing for fewer essential staff. The price from Bepress appears to be at significant discount, especially compared to other consortial pricing (cf. Bepress offer to CSU system in 2013). There is no local repository manager, no metadata specialist, and no SCA position associated with the IR. The campus coordinator contributions vary greatly, roughly between five hours a week and five hours a semester.

Lessons / takeaways

Examining CUNY's repository, it seems possible to create a repository for a consortium as large and as varied as the CSU (in terms of size, stakeholders, and users) using just one person as a coordinator. As we all know, Bepress and the CSU CO were in talks five years ago. The quote given to the CSU at the time was approximately \$1 million, not including long-term storage costs or data transfer fees.

Looking at CUNY's collections, however, one can see the limitations of their approach. Only **19,000** items are archived in this collection among their 25 campuses and 31 libraries. In comparison, the CSU system has archived well over **200,000** items in both ScholarWorks and

local campus repositories (**See Appendix B**). This approach will therefore likely be counterproductive given the large collections already in ScholarWorks, as well as antithetical to the COLD resolution on implementing open source solutions.

Mountain Scholar: Digital Collections of Colorado and Wyoming

- Platform: DSpace (IR) and Kaltura (DL) streaming server
- Platform purpose: Digital Collections and Institutional Repository.
- Collections scope: Scopes vary from community to community but the overall repository contains faculty research, ETDs, archival collections, journals, technical reports and datasets.
- Collection size: 89,430

Brief

Mountain Scholar includes **nine** public colleges and universities located in both Colorado and Wyoming. All are employees of Colorado State University.

Background and governance / funding

Colorado State University is also the holder of an NSF grant for the Rocky Mountain Advanced Computing Consortium, which is currently covering much of the infrastructure. The repository is also supported by each participating institution **paying \$8,000** annually, which would amount to approximately **\$72,000/year**. Since the funds they received from the state will eventually run out, Mountain Scholar is considering raising the membership fees to offset the loss of funding.

Lessons/takeaways:

These are exactly the same platforms the CSU system has used in the past. A point of distinction from the way our system has implemented these platforms is that one university has taken the primary lead in terms of providing the base service for the consortium. The yearly fees for membership appear to be an acceptable and reasonable cost for members. Their overall collection size of nearly **90,000 items** includes both digital collections and repository materials.

Oregon Digital (UO and OSU)

- Platform: DSpace *and* Samvera
- Platform purpose: More of a suite of services, somewhat comparable to Texas Digital. System, once revised (an ongoing process as of now) will accommodate both IR and unique collections.
- Collections scope: Cultural heritage materials only
- Collection size: 252,810

Brief:

Oregon Digital is a joint collection of unique digitized and born-digital materials including photographs, articles, sheet music, manuscripts, ephemera, and more, held by the University of Oregon and Oregon State University. Both UO and OSU have separate IRs; UO is hosted on DSpace and OSU recently migrated to a Hyrax. Oregon Digital itself is undergoing a “re-write” from top to bottom in order to better support IR and digital collections needs. The initial build was created to accommodate a migration from CONTENTdm, not necessarily for some of the particularities of working with ETDs.

Background and governance / funding

Overall costs are unclear, but it seems that each university supports the staffing costs for their respective campuses and repositories. The staffing includes the equivalent of about 3.6 FTE, so perhaps about **\$324,000 (@ \$90k)** per year? The University of Oregon’s IR on its own hosts the following staffing structure: Institutional Repository Manager, Head of Digital Scholarship Services, Programmer, Director of Library Technology Services, Assistant Director of Library Systems, Analyst Programmer, Supervisor of Library Applications Programming and a Systems Administrator.

Lessons / takeaways

The CO has worked with this group previously to develop migration scripts from DSpace to Hyrax. Oregon Digital project lead, Steve Van Tuyl, at Oregon State University, is also currently

the product owner for Hyrax in the Samvera community. The concerning thing about their model is the split between DSpace and Hyrax for the two campuses, and the issues of overlap in personnel working on the projects (i.e. developers, metadata experts, outreach specialists), though we are not certain if this model is going to continue this way permanently, or if both IRs will be integrated into the new Oregon Digital build.

State University of New York (SUNY)

- Platform: DSpace
- Platform purpose: Digital Collections and Institutional Repository.
- Collections scope: Acts more as a catch-all digital repository in that there are no parameters limiting its scope. There are multiple contributors (consortial) and some have photographs and digitized archival materials along with research output (such as ETDs, faculty publications, etc.)
- Collections Size: 25,490

Brief:

SUNYConnect includes 39 SUNY Colleges and Universities, Alfred College of Ceramics, an OER Repository, SUNY Strategic Planning, and SUNY Administration collections.

Background and governance / funding

Unclear. As of drafting this report, we did not receive details on this topic from SUNY.

Lessons / takeaways

SUNY is the largest university system in the United States. As number two, the CSU is frequently compared to SUNY in discussions of large university systems. Since SUNY also uses DSpace, we imagine that we could gain some insight into their strategies for sustainability and funding. Their approaches would likely provide a few lessons for the CSU. However, looking at their collection size (only **25,490 items**) relative to the number of organizations involved (nearly 40),

it is likely an extremely underfunded and understaffed endeavor. For a notable comparison, **San Jose has over 25,000** and **Northridge has over 30,000** items respectively.

Texas Digital Library (TDL)

- Platform: DSpace
- Platform purpose: Digital Collections and Institutional Repository. Acts as a digital repository for “unique collections of enduring value” for member institutions.
- Collections scope: Provides digital preservation, IR needs, data repository, and a forum for open access publishing.
- Collections size: 57,366

Brief:

The TDL began in 2005 as a partnership between four of the state’s largest Association of Research Libraries (ARL) universities. It now includes **22** members (colleges and universities across all of Texas). TDL employs eight people to run the various repositories for its members.

Background and governance / funding

The TDL membership is represented by a Member Board, which includes the administrative head of each Regular Member (plus a representative administrative head for each Consortia Member). The Member Board meets yearly in the fall to discuss issues of concern to the membership and to elect at-large members of the Governing Board. The TDL Governing Board provides strategic direction for the Texas Digital Library and is comprised of library deans and directors from seven TDL institutions. The founding ARL members of TDL serve as ex officio members of the Governing Board. The remaining three members are elected at-large from within the regular membership. In addition, two members of the TDL staff sit on the Governing Board as ex officio, non-voting members.

Each institution involved with the TDL pays for services based on the Carnegie designation of its campus. The tiered model is as follows:

ARL Institutions	Doctoral Universities: Highest Research Activity	\$75,000
	Texas A&M University	
	Texas Tech University	
	University of Texas at Austin	
	University of Houston	
High Research Activity Institutions	All other Doctoral Universities and Special Focus Institutions	\$20,000
	Public and Private Institutions	
	Four-year Medical Schools and Centers	
Master's Colleges and Universities	Master's Colleges and Universities (also includes branch campuses of flagship universities)	\$10,000
Private Liberal Arts Colleges and Universities	Baccalaureate Colleges and Universities	\$5,000
Community Colleges	Community Colleges	\$2,000

Additionally, TDL provides specific services for each campus should they desire it, ranging from DSpace hosting, Thesis submission management system (Vireo) hosting, journal hosting through OJS, digital preservation, and access to the Texas Data Repository.

Lessons / takeaways

The tiered fee model for funding the consortium might make sense for the CSU, though more in terms of campus size rather than Carnegie designations. Partnering with other colleges and universities outside the CSU system may be a good future goal to consider. The TDL service module pricing also seems like a good cost-effective approach to providing services based on specific campus needs. We could model a lot of our structure on the TDL and wind up with a very successful repository.

Washington Research Libraries Consortium (WRLC) Digital Collections

- Platform: Islandora
- Platform purpose: Digital Collections and Institutional Repository. IR/Data hosting and sharing/DL all treated as distinct services.
- Collections scope: ETDs, faculty research, datasets, archival collections.
- Collections size: 105,850 items

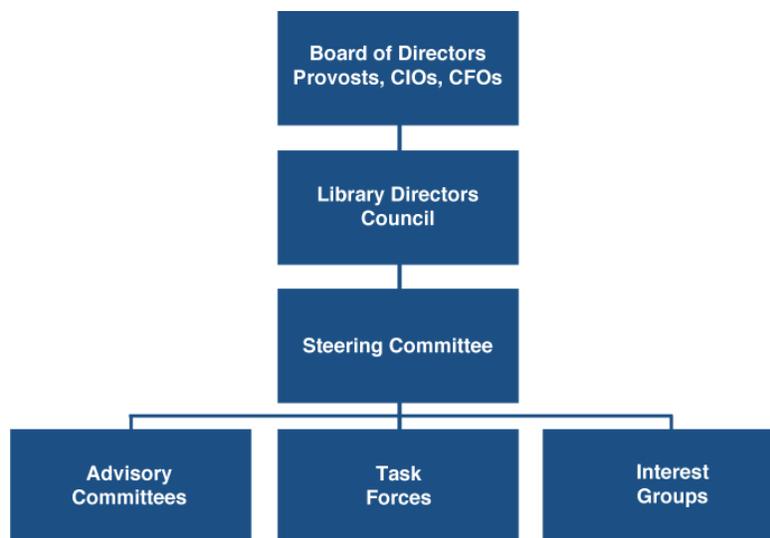
Brief:

Washington Research Library Consortium provides IR, data hosting, and digital collections for nine private universities in the Washington D.C. area.

Background and governance / funding

There are three positions that directly support the IR, which are centrally funded. Only 10% of their time is actually spent on the IR. The cost of these is approximately \$20,790. Additionally, outsourced support for Islandora is \$20,000 / year shared among the nine member institutions. The total cost for the repository / DL is approximately \$40,000 per year.

The governance model for WRLC is as such:



Lessons / takeaways

This is a very robust and long-standing cooperative model. The consortium itself has existed since 1987 and as a result seems to have a well-established set of protocols and governance structures in place to handle a wide variety of services, including repository/digital collections. The costs seem minor in comparison to Bepress or other services, while the overall size of collections seems fairly large given the size of the institutions involved and their Carnegie

classifications. The CO has had several discussions with WRLC regarding their project over the years.

OhioLink

- Platform: OhioLINK ETD Center, homegrown using an Oracle database
- Platform purpose: OhioLINK's ETD Center is an ETD service only
- Collections scope: ETDs
- Collections size: 96,332

Brief:

The OhioLINK ETD Center is a homegrown system, created and run by OhioLINK and available to any of the OhioLINK members who want to use it. Reviewing ETDs is left up to each institution, with their own submissions and set local policies and workflows.

Background and governance / funding

The ETD Center staffing model includes a Metadata and ETD Coordinator, a Manager of Digital Platforms, and some in-house developers (various titles). Thirty-three out of 118 OhioLINK members participate in the ETD Center. OhioLINK itself is a resource sharing model and the ETD Center is a membership benefit. This involves using the system, voting on enhancements, testing updates and offering feedback. According to the OhioLINK respondent, "Some members use the ETD Center year-round for all of their submissions while some only have certain departments submit, and still others only upload a handful of stellar ETDs (very small schools) once a year to represent work done on campus; the level of involvement with the system for local use is entirely up to each institution and what fits their needs and sometimes staffing levels." The ETD Center is a small piece of a bigger digital platforms group therefore the overall costs are hard to determine. There are no vendor costs, as the system is homegrown.

Lessons / takeaways

The respondent from OhioLINK believes the discoverability of the ETDs is improved by being in a single system. All participating institutions are in one place instead of spread out in local IRs. Also, Google crawls the ETD Center so they are all easily findable online and included in Google Scholar. Troubleshooting is also made easier with everyone in the same system. The process is further streamlined and solutions are reused.

MOSpace (No longer a centralized repository)

- Platform: DSpace (IR), Islandora (DL)
- Platform purpose: Digital Collections and Institutional Repository
- Collections scope: ETDs, faculty and student research, a/v materials, pre-prints, published materials, digital art, post-prints, manuscripts, journals and more.
- Collections size: N/A

Brief:

MOSpace was the centralized institutional repository for five University of Missouri campuses. After centralizing, two campuses determined that DSpace did not meet their needs and moved to Digital Commons. Three campuses continue to share Islandora for their digital library collections.

Background and governance / funding

While centralized, the University of Missouri campuses shared technology staff, but each campus maintained its own staff for outreach, submissions, and cataloging. The campuses did create teams to coordinate workflow and metadata. The respondent did not provide the specifics of FTE dedicated to the project.

Lessons / takeaways

We thought it important to continue to include MOSpace, even though the centralized IR was not successful, because the respondent provided some benefits and challenges of their

experience with a centralized system. The consortia found the shared technology staff and collaborative inter-campus teams to be beneficial for developing workflows and maximizing development. They also believe a single repository made it easier to highlight the contributions of the University of Missouri as a system. However, the experience of shared indexes and metadata standards was challenging and the negotiation process, while useful, was time consuming. They did eventually create special indexes to meet the needs of different campuses, but that does create an additional maintenance burden for the system. The respondent doesn't believe that a single instance necessarily makes content more discoverable, since most users come to MOSpace via Google searches. They did provide, however, that a single instance could lead to exploration and additional discovery of system-wide materials once users were at the repository site.

Conclusions

Looking at each of these consortial models provides us with a lot of separate approaches to building a system-wide repository. Each of these models uses different platforms, ranging from a home-grown, tailor-made system to Bepress (a fully 3rd party hosted solution) to well-established open source solutions such as DSpace, Islandora, and Hyrax. The lack of a uniform solution or funding model points to the different circumstances for each of the institutions, their constituents, their stakeholders and their users.

The successful repositories, despite the lack of uniformity in their platforms, have notable similarities: mainly an emphasis on both consistent governance and infrastructure (both technical and funding infrastructure support). Defined policies and best practices, along with technical and financial transparency, seem to allow for more sustainable project management. This suggests that the platform may be the *least* important aspect of a repository. A platform is just a platform, after all. 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. This is something the CSU system must grapple with in order to become more successful. Yet the good news is that we are nevertheless a growing consortium in our own right, with over 200,000 digital records across 23 campuses and multiple repository platforms, ranging from a few hundred items to over 30,000. We are poised to grow, but how we do so may be dependent in part upon what aspects of these different consortial models we choose to adopt. A hybrid approach may serve us best.

APPENDIX A: Questions posed to participating consortia

- 1) Please provide the titles of staff members that directly support the IR
- 2) What is the level of required involvement from each participating institution
- 3) Which platform are you using? If you are using a proprietary platform, are you considering moving to open source system and, if so, which one?
- 4) What is the total cost of operation (a ballpark figure is fine if you do not have the exact cost)

Please break out the cost by the following, if possible and appropriate:

- centrally funded positions
- vendor support
- other outsourced support
- other costs

- 5) What are the individual costs to the institutions, if cost is shared (average is fine)
- 6) Is there anything else you would like to share regarding the value of one system for all? And, specifically, do you feel this has been a successful configuration for your institutions?

APPENDIX B: CSU Institutional Repository Collections Totals

Campus	Total objects
Bakersfield	2,039
Channel Islands	17,636
Chico	939
Dominguez Hills	0
East Bay	745
Fresno	12,252
Fullerton	226
Humboldt	3,061
Los Angeles	4,381
Long Beach	4,786
Maritime	5,300
Monterey Bay	7,332
Moss Landing	267
Northridge	30,533
Pomona	2,538
Sacramento	4,306
San Bernardino	11,718
San Diego	21,644
San Francisco	3,856
San Jose	25,459
San Luis Obispo	34,272
San Marcos	3,048
Sonoma	3,138
Stanislaus	642
Academic Tech. projects	16,495
Total	216,613

Report 4: Technical Considerations for a Shared Institutional Repository

The need for multitenancy

A core technical requirement for the ScholarWorks project is multi-tenancy – that is, a deployment strategy in which components of an application are shared by all campuses. Implementing and supporting 23 fully separate installations of any complex system is enormously (and needlessly) costly. For that reason, all of our existing system-wide applications – including Alma, Primo, and DSpace – utilize some kind of multi-tenancy.

The digital library platform we are migrating to, Hyrax, does not currently support multi-tenancy out-of-the-box. Although the Samvera community has begun to address this need through a project called Hyku, that work is still in-development at the time of this report. To date, the Chancellor’s Office has needed to employ a different strategy: separate instances of Hyrax for each campus with shared back-end data storage (Fedora) and indexing (Solr) components.

Although this type of loose multi-tenant configuration would give each campus much greater flexibility over what we have today with our multi-tenant DSpace installation, it comes at a pretty significant cost, requiring separate servers, interface customizations, and configurations for each campus. Amazon Web Services and other software tools (e.g., Git) give us some means to ease the duplicate work involved here, but it is nevertheless a large and complex deployment to set-up and maintain.

Demo system

The SWAT team explored an alternative strategy: a single instance of Hyrax shared by all campuses. The [demo system](#) was designed to explore the feasibility of such an approach, by

examining how a single-instance repository might support (1) authentication, (2) the ability to submit, approve, and assign works to a specific campus, (3) the ability to have separate controlled vocabularies per campus, specifically to support unique college and department names, and (4) embargoes.

1. Authentication. In the production system, when a user selects the login option in the menu, they will be directed to the Chancellor's Office's 'discovery' or WAYF (Where Are You From?) service, which will allow the user to select their home campus and login to their campus single-sign-on service (e.g., Shibboleth). This is how CSYou and other system-wide sites work today. In the demo, we just created local accounts to mimic this behavior, as the technical requirements for this are already well understood.

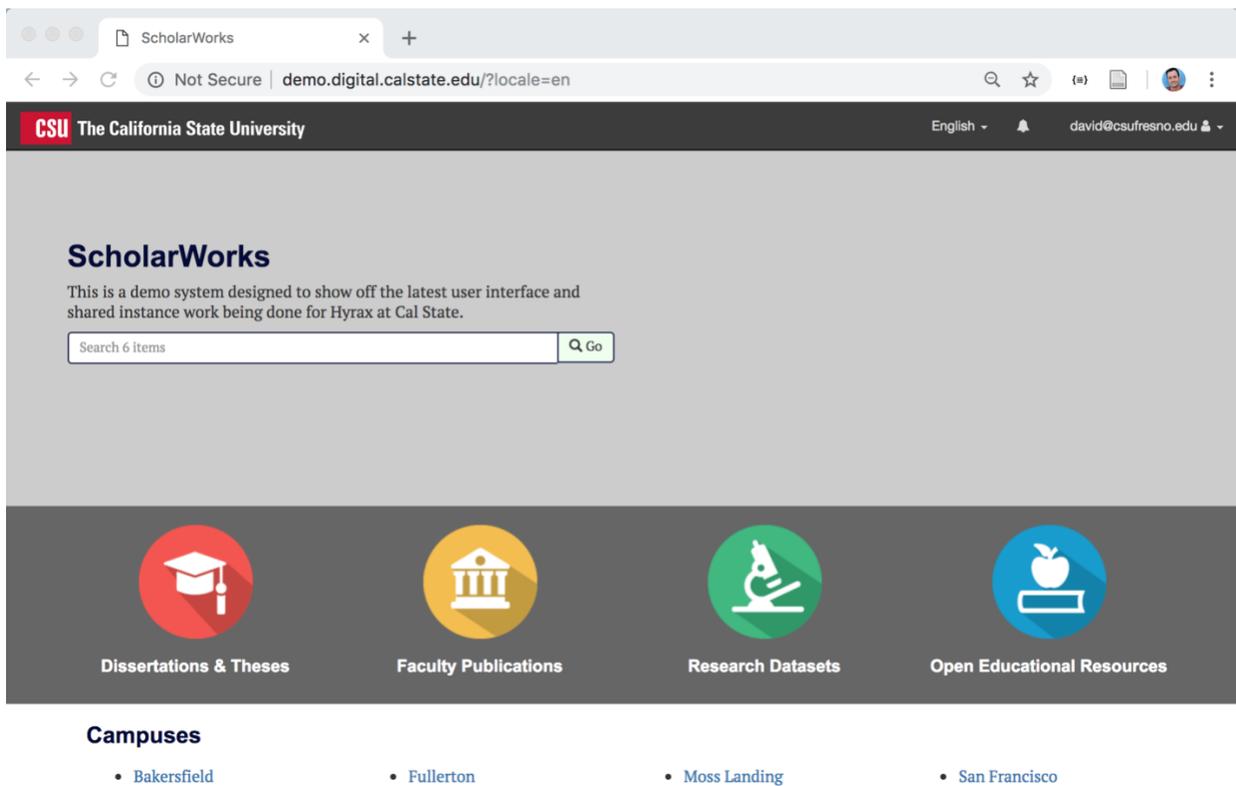
2. Campus-specific submissions and approvals. Once the user has been authenticated, Hyrax will map them to a group based on their campus affiliation. A student at Fresno, for example, will be assigned to the 'fresno' group, a faculty member at San Marcos to the 'sanmarcos' group, and so on. The groups themselves will be assigned as 'submitters' to one or more administrative sets in Hyrax, which will allow campuses to select different workflows for different types of material – e.g., a mediated workflow for theses, an unmediated workflow for faculty publications, and so on. Hyrax will automatically assign the campus name to each submission, based on the submitter's group. This will ultimately allow end-users to limit their search to results from a specific campus.

3. Campus-specific controlled vocabularies. The demo system includes a customization that allows each campus to specify its own controlled vocabulary for any field. This is specifically designed for the case of colleges and department, which vary from campus to campus, and are often ascribed to ETDs and faculty publications.

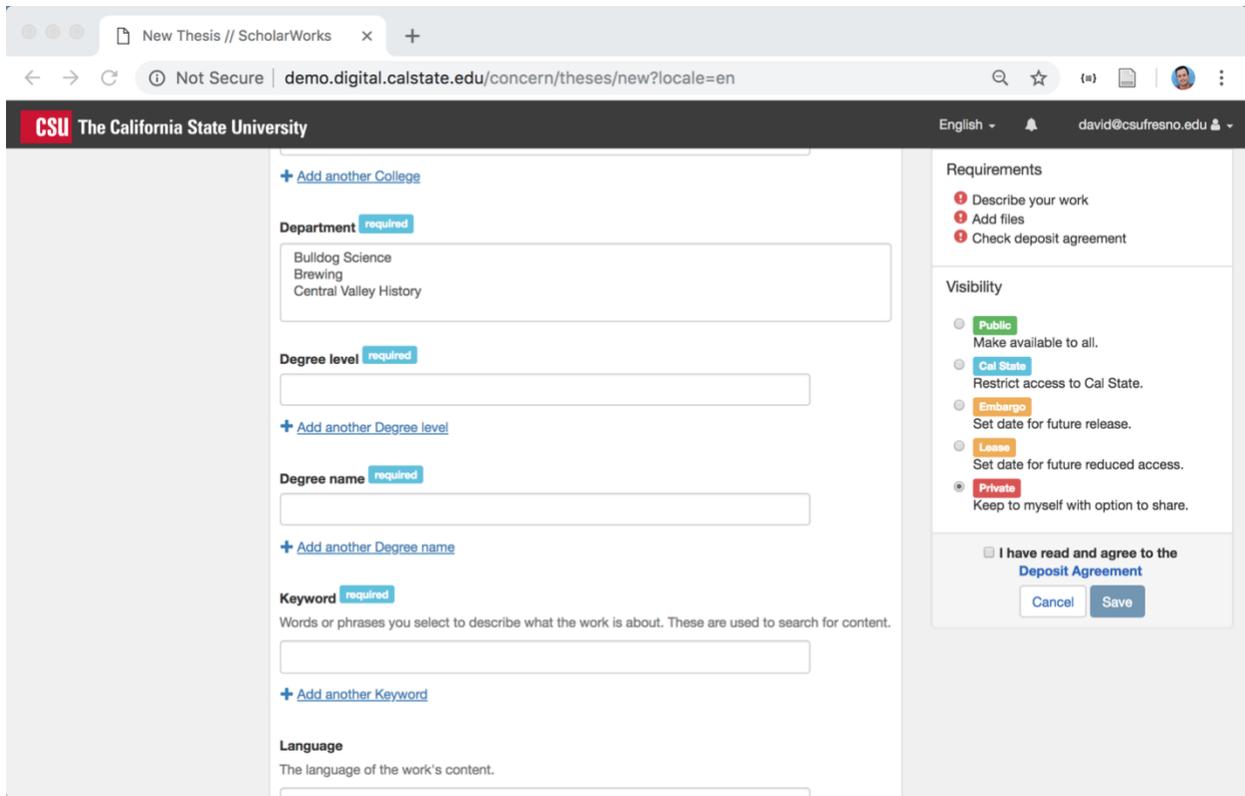
4. Embargos. Hyrax natively supports two different visibility options for embargoed items: 'private', which allows only the submitter to view the item, and 'institution', which allows logged in users to view the item. In a shared instance, such as with the demo, the 'institution' would include all authenticated CSU users, and not just users at a specific campus. Hyrax does

support the option of granting a specific group access to view an item, however, and so we could create a campus-only visibility for embargos by assigning 'private' visibility and giving the campus group read access. This would need to be a customization.

On the whole, the demo system also illustrates the 'flatter' nature of Hyrax's approach to metadata, especially compared to DSpace with its nested communities and collections.



Screen shot of home page showing Fresno user logged in. End-users can search and browse by broad content types (these are merely examples, actual categories TBD) or by campus.



Demo showing Fresno-specific departments for a Fresno user submitting a thesis.