ScholarWorks Planning Advisory Report

February, 2016

Purpose

The *ScholarWorks* Planning & Advisory group has reconvened the available members of the 2013 STIM Subcommittee in order to assess the progress of the CSU Institutional Repository projects. Additional members were added to review and act as additional pilot users for new projects.

Membership

Aaron Collier, Chancellor's Office
Andrew Weiss, Northridge
Elizabeth Altman, Northridge
Carmen Mitchell, San Marcos
Ian Chan, San Marcos
Suzanna Conrad, Pomona
Jeremy Shellhase, Humboldt
Emily Chan, San Jose
Jayati Chaudhuri, Los Angeles

2016-2017 *ScholarWorks* Roadmap

1. Adopt *Fedora Commons* as the digital repository platform used by *ScholarWorks* (draft migration project plan to be attached)
2. Continue to use the current Amazon Glacier system that is in place for digital preservation.
3. Offer an additional suite of services as part of of *ScholarWorks*:
    a. Faculty Profiles (through *Islandora*/*IslandScholar*)
    b. Publishing support (via *Open Journal Systems* /*Open Monograph Press*) – *OJS* currently in BETA
    c. Conference/Symposium support (via *Open Conference Systems*)
4. Establish an annual *ScholarWorks* symposium with alternate hosting between northern and southern campuses to facilitate ongoing in-person training and outreach opportunities.
5. Develop the *ScholarWorks* confluence site as the up-to-date communication mechanism for news, documentation, changes, and status updates related to *ScholarWorks* projects.
    a. Compile existing workflow documentation, processes, and outreach materials from each campus within confluence.
    b. Utilize regular IR meetings to plan workshops on outreach and workflow, as needed.
6. Adopt a formal customer support platform (i.e. ticketing and issue tracking system) to provide effective ongoing customer support to all users of hosted *ScholarWorks* projects.

Executive Summary

This report is divided into the following sections:

I. Update to STIM IR Subcommittee Report from April 2013
II. Status update on *ScholarWorks* since 2013 report
III. Review and justification for adopting *Fedora Commons*
IV. Recommendation of services that should be added to the suite of *ScholarWorks*
V. Outreach, training and support

In April 2013, The STIM Institutional Repository Subcommittee delivered a report on the state of *ScholarWorks*. Ultimately the recommendation of that report was to continue the use of *DSpace* as the core technology platform used for all *ScholarWorks* hosted CSU institutional Repositories. Additionally, the subcommittee requested that a task force continue the review the evolution of open source repository platforms that were not feasible at the time of the report. That request was for another review by the summer of 2016. This report addresses that request. Section I of this report will provide a brief update on the core features of that report.

A substantial theme of the 2013 report was the total cost of ownership (TCO) comparison between the Chancellor's Office hosted *DSpace* platform and the commercial platform *Digital Commons*. That review was very thorough, and primarily has not changed in the time since. For that reason, this report will not extensively address TCO, except where appropriate updates may be pertinent. It is worth noting that some CSU campuses have independently moved forward with licensing *Digital Commons,* even though this was not the recommended action. This committee is interested in learning the cumulative annual subscription amounts that CSU campuses are spending on *Digital Commons*.

In section II, a status report of projects and updates to the *DSpace* platform provided by the Chancellor's Office are outlined.

Section III reviews the benefits offered by *Fedora Commons* as a core digital repository technology platform and the display platform technologies that are exposed by adopting *Fedora*. A feature comparison between *Fedora*, *DSpace*, and display platforms where appropriate is provided. *Digital Commons* is not included in this review as the pricing has not changed and remains a cost prohibitive solution. While the adoption of any core repository technology will affect all hosted *ScholarWorks* campuses, we feel this recommendation is critical to the continued success of the CSU digital repositories as a whole.

Section IV outlines several recommended services that functionally relate to ScholarWorks. These services include:

- A hosted publishing platform for journals, monographs and conference/event organization
- Faculty profiles
- Formal customer service portal

Finally, Section V presents a recommendation to develop an annual CSU IR symposium that will provide in-person training, collaboration and coordination opportunities for existing and new IR staff members today and going forward.

Analysis and Discussion

Section I: 2013 STIM Institutional Repository Subcommittee Report Update

The bulk of the 2013 report was focused on a "Total Cost of Ownership Analysis" between *DSpace* and *Digital Commons*. This has largely remained unchanged in the time since the report was presented, however, changes to the infrastructure and responsibilities at the Chancellor's Office have further improved the TCO of the hosted repository platform. Notably this includes:

- Support tasks are being managed by Digital Repository Services Manager Aaron Collier, allowing the contracting costs to be lowered by $10,000.
- Chancellor's Office IT infrastructure upgrades allowed the migration of the cloud storage off of the *AtMos* appliance, saving $10,000 in licensing and maintenance fees.

It is the recommendation of this group that the CO-hosted model for IR related projects be continued and expanded where possible. Additionally, we seek a position that, under the direction of Aaron Collier, would advance *ScholarWorks* by developing and documenting best practices, writing documentation, developing marketing materials, clarifying copyright issues, and coordinating open access initiatives and collaborative efforts across campuses[1]. Having a centralized position would standardize practices, more efficiently use limited resources, and decrease repetitive work and development currently occurring in libraries across the CSU system.

Additional recommendations from the 2013 report and status are:

- Examine open source solutions for journal publishing and other IR services – An instance of *Open Journal Systems (OJS)* has been installed. The first publishing project to utilize this platform will be started during the spring of 2016.
- Begin investigation of next-generation open source IR platforms – At the time of the 2013 report, *Fedora Commons* (using the *Islandora* or *Hydra* interface) was very new and presented an unknown landscape for its future direction. Since that time, it has developed into a robust and widely-used institutional repository platform.
- Audio and Video Streaming – at the time of the report, the Chancellor's Office was working to implement *Kaltura* integration into *DSpace* in order to provide streaming support. It was later determined that the *Kaltura* platform was incompatible with the *AtMos* storage appliance. Due to that finding, the Chancellor's Office utilized existing support credit with @Mire to integrate a cloud based streaming platform, hosted on Amazon Web Services (AWS), into *DSpace* at no additional cost. This project was completed during the summer of 2015 and has proven to be very reliable. It will be the recommendation of this report that any new IR platform migration include this streaming platform as a requirement.

[1] http://osc.universityofcalifornia.edu/open-access-policy/

Section II: *ScholarWorks* Status Update

Utilization of Chancellor's Office hosted Institutional Repository

Since the 2013 report, two additional campuses (Cal Poly Pomona and CSU Stanislaus) have adopted the hosted *DSpace* instance provided by the Chancellor's Office. Three non-campus specific projects are underway utilizing the hosted repository: the Quality Online Learning and Teaching (QuARRy) Resource Repository, the COAST faculty repository, and the COOL4Ed repository. Additionally, the CSU Visual Collective is being integrated into the core *DSpace* instance in order to provide a more reliable platform for that project.

ScholarWorks Development Projects

The most notable project completed within the *ScholarWorks* platform was of consolidation of *DSpace* from separate installations for each campus to the current multitenant structure. When each campus had its own "instance" installed, any upgrades or patches had to be done on each separate instance. It was time consuming, inefficient, and costly to maintain. By migrating to multitenancy, *ScholarWorks* has become more stable, upgrades are simpler to plan and implement, and downtime has been minimized. The hiring of the Digital Repository Services Manager made this migration possible. A batch item submission portal was deployed in early 2014. This portal utilizes the SWORD (Simple Web-service Offering Repository Deposit) protocol, allowing continued use going forward regardless of repository platform. SWORD made several projects throughout the CSU repositories possible through easy data entry, saving many hours of duplicate work. CSU San Marcos has uploaded over 900 metadata records for their retro-thesis digitization project, Cal Poly Pomona frequently uses this tool to upload batch content from conferences, on-campus journals, and undergraduate research projects, CSU Northridge has utilized the portal for retro-thesis metadata overlay and batch uploading of over 2400 items to date. This project is maintained outside of *DSpace* and will continue to be available, regardless of repository platform.

A statistics portal was developed that provides a more accurate view into system usage than is available by default with *DSpace*.

As noted above, a cloud based multimedia streaming service was integrated into *DSpace*.

Section III: Review and Justification for Adopting Fedora Commons

Some historical background: CSU Adopted *Dspace* as the repository platform summer of 2007. At that time, while *Fedora* was firmly in version 2.0, *Islandora* was very new and deployment of the technology stack required developer time and skill. It was far from turn key. Since that time, *Islandora* has grown significantly into a reputable digital repository platform, *Hydra* has grown substantially and is approaching a turn key installation, however *DSpace* development has largely stagnated. The community for *Hydra* development includes Stanford, Penn State, Duke University, Yale University, and the Digital Public Library of America, and many others.[2] *DSpace* has an outdated look and feel. It also lacks a built-in RESTful API. This means that integrating *DSpace* content into other systems is difficult, interoperability is difficult, and it is not optimized for linked data. *DSpace* does great with

[2] https://wiki.duraspace.org/display/hydra/Partners+and+Implementations

text-based materials like theses and journal articles, but is not optimized for images. *Hydra* utilizes Ruby on Rails, includes a REST API, and is geared towards linked data. Like *Hydra, Islandora* utilizes a *Fedora* repository back end while being deployed through the very well established *Drupal* content management system. The community surrounding *Drupal* development is very active, allowing plug and play solutions to almost an feature request possible.

Notable examples of *Islandora* Repositories:

Moss Landing Marine Laboratories: http://digital.mlml.calstate.edu/

Florida State University Library: http://diginole.lib.fsu.edu/

Michigan State University Electronic Theses & Dissertations: http://etd.lib.msu.edu/

University of Prince Edward Island IslandScholar: http://www.islandscholar.ca/

Notable examples of *Hydra* Repositories:

University of Hull: https://hydra.hull.ac.uk/

University of Virginia: http://libra.virginia.edu/

Penn State: https://scholarsphere.psu.edu/

Stanford University: https://exhibits.stanford.edu/

Section IV: Service Additions to *ScholarWorks*

Publishing - Journals and Monographs

There has been a lot of interest in launching journals throughout the CSU, as well as maintaining existing print journals that need a reliable digital platform. Including a journal publishing platform (via *OJS* - Open Journal Systems) in the *ScholarWorks* repository suite makes sense in order to coordinate the communication between publishing, streamline the workflow process, and facilitate repository archiving following current digital preservation standards and best practices.

Faculty Profile Management

In addition to capturing the scholarly output of each campus's faculty, there is added difficulty in maintaining the status of faculty citations, research areas of interest, etc. Some campuses have expressed a need for improved functions for featuring faculty as a means of establishing subject area experts for media inquiries, improving assessment and tracking of scholarly endeavours at the college level, and alleviating some webspace requirements for IT departments by providing an alternate online space for faculty accomplishments. Recent *Islandora* developments, especially those implemented by the Library at the University of Prince Edward Island, have shown a robust user interface for presenting faculty profiles. Additionally, *Islandora* has solution packs for managing faculty scholarship including integration with Sherpa/RoMEO, options for citation batch ingests from common citation programs, and interlinking with Google Scholar.

Section V: *ScholarWorks* Outreach, Training and Support

From the 2013 STIM Report, "The data collected appears to indicate that regardless of the platform used, the success of the IR largely depends on the local campus's commitment to staffing the IR."

As staffing of repository projects grows, outreach, documentation, help-desk level support and project communication must be maintained centrally in order to provide a consistent level of service without impacting ongoing development efforts.

Outreach

The 2015 CSU *ScholarWorks* Conference successfully brought together repository librarians and staff from across the CSU to report on the completion of the multitenant migration project, review and discuss the 2013 STIM IR Subcommittee report and present on local campus projects within the repository. Katie Fortney, the Copyright Policy & Education Officer for the University of California Office of Scholarly Communication was the guest speaker, presenting on how Symplectic Elements has automated the deposit and ingest of UC faculty works into eScholarship, the UC institutional repository.

This day-long meeting was incredibly successful in bringing the repository related personnel from across the CSU together in order to openly discuss the status and future direction of ScholarWorks. By expanding this format to follow the very well respected University of San Diego Digital Initiatives Symposium (http://digital.sandiego.edu/symposium/2016/), a half day of presentations training sessions could be added to the schedule for new staff members as well as refresher courses for existing users. Additionally, if this symposium could act as a catalyst for engaging the CSU Library IT group in more frequent in person communication, it would afford technical staff an opportunity to collaborate on the more technical needs of Scholarworks through hackathons and other co-development strategies.

In-person communication, especially when focused on training, is generally superior in efficiency and effectiveness than virtual methods of communication. However, with stakeholders at all 23 campuses of the CSU, and limitations of travel budgets and time, in-person communication must be collaborative, well organized and planned far in advance. A scheduled, annual symposium, focused on ScholarWorks support and training will provide the best possible platform for outreach with sufficient time for individual planning. By alternating locations between northern and southern California, the impact on travel for all campus users will be significantly reduced.

Documentation

The CSU Chancellor's Office adopted the Atlassian Confluence platform to capture and communicate documentation as part of the ULMS project. Using the excellent example of format and organization for that project, the *ScholarWorks* Confluence pages (https://calstate.atlassian.net/wiki/display/SCHOL) will continue to expand in order to centralize documentation for all related projects.

Support platform

As @Mire was brought into the support structure of *ScholarWorks* prior to the addition of Aaron Collier to the project, utilizing their internal ticket tracking system has been sufficient for project and

support monitoring required for the *DSpace* platform. This system will not be available to us as a support platform when moving off of *DSpace*.

Atlassian's JIRA product is available and integrated into Confluence and would make the most sense to adopt as a new support platform. A pricing investigation is under way. JIRA offers both a help desk platform and a project management platform that with integration into the existing confluence site will greatly enhance the communication capabilities during this project.

Appendices

Appendix 1: Proposed Migration Timeline

Phase 1 **Information Gathering, Interface Selection, and Data Migration Planning**			
Task	Start Date	End Date	Notes/Status
Fedora test & dev installation	1/7/16	1/7/16	Current version is 4.4.0, Current Islandora and Hydra supported version (as of this report) is 3.8.1 COMPLETE *migration planning to 4.x to be included
Data migration development	1/13/16	2/10/16	Initial data migration development will focus on the translation of *DSpace* exported items to a supported *Fedora* import format *additional migration planning is included in phase 2
Islandora demo installation	1/11/16	1/12/16	
Hydra demo installation	2/1/16	2/2/16	
Survey repository librarians, select users, and other interested parties	2/8/16	2/29/16	
Survey data coding and requirements	2/29/16	3/21/16	

outline development			
Interface testing against requirements	3/21/16	5/2/16	
Proposed interface selection	5/6/16	5/6/16	
Phase 2: Data migration planning & testing, wave 1 campus selection, and demo system installation			
Phase 3: Wave 1 campus data migration, verification, and go live			
Phase 4: Wave 2 campus scheduling and decommissioning of *DSpace*.			

Appendix 2: Preliminary Platform Migration System Requirements

1. Must maintain existing archival storage in Amazon Glacier.
2. Must maintain streaming media support.
3. Must support Handle registry.
4. Must support linked data.
5. Must support embargos and other forms of access restriction.
6. Should allow SWORD Support for CSUN application and bulk item submission.
7. Should allow for individual collection theme application.
8. Should allow for collection-level submission form templates.
9. Should allow for individual item type submission workflow.
10. Should allow integrated faculty profiles.
11. Should allow for batch import and export of records by non-programmers.
12. Should allow for more granular permissions than at present (e.g. optional contributor access to correct/remove/replace records and files after submitting; option to authorize catalogers to export/import records through the native interface without having to be collection manager; optional contributor access to statistical reports).

Note: The migration requirements will be updated through survey feedback and testing.

Appendix 3: Institutional Repository Platform Review Summary

Disclaimer: The final implementation plan is subject to change.

Choice	Platform	Pros/Cons
First	*Fedora*	Pro: Single platform solution for digital object and metadata storage. Clean UI and better display options for images and non-textual items. With *Hydra*, customizations can more easily done than with *DSpace*. Currently supported by a robust community and code-committer base.
Second	*DSpace*	Pro: Current system, already functional. Con: requires multi-tier installation of both a database AND SOLR for metadata storage. UI is not easy to use. Customizations are difficult. Administration requires a lot of training. Not in active development/no rapid innovations happening.
Third	*Digital Commons*	Pro: Hosted solution. Nice UI for administrators and end-users. Cons: Expensive. Requires additional preservation plan OR additional subscription for DC preservation plan. Places CSU-created content into a platform hosted by a for-profit company. Any migration off of DC would incur additional expense. Limited system integration options

Appendix 4: 2015 ScholarWorks Conference Report to COLD

Group Consensus Document from the CSU Scholarworks Conference April 10, 2015, CSU Scholarworks Conference, San Jose State University

The following statement was unanimously agreed upon by all attendees at the end of the conference:

In order to increase the profile of the CSU, to augment the amount of content curated in Institutional Repositories (ScholarWorks, Digital Commons, etc.), and to better communicate their usage we ask COLD to develop and fund an additional position to focus the work of the 23 campuses in the development and use of the Institutional Repositories. Following the recommendations from the STIM IR Subcommittee report of 2013, we seek a position that, under the direction of David Walker, would advance IRs by developing and documenting best practices, writing documentation, developing marketing materials, clarifying copyright issues, and coordinating open access initiatives and collaborative efforts across campuses. Having a centralized position would standardize practices, more efficiently use limited resources, and decrease repetitive work and development currently occurring in libraries across the CSU system.