CSU Digital Archives DAMS Pilot

A Report of the Digital Archives Working Group

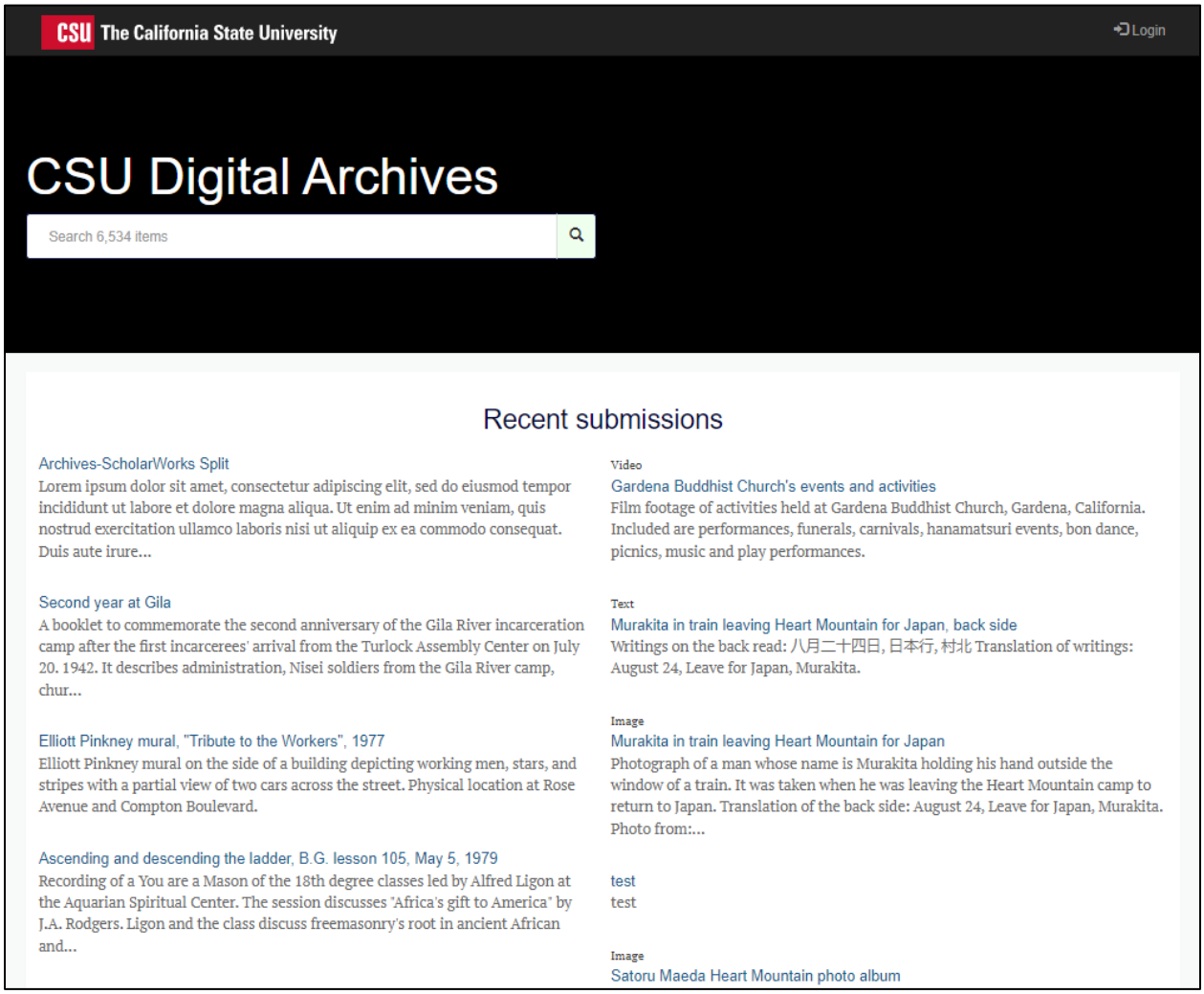

Image: Home page from CSU Archives pilot installation.

Steve Kutay, California State University, Northridge
February 10, 2023

Contents

Acknowledgements

This assessment was made possible by generous contributions from members of the CSU community:

- Elizabeth Blackwood, CSUCI
- Julie Dinkins, Sonoma State
- Steve Kutay, CSUN
- Carmen Mitchell, CSUSM
- Yoko Okunishi, CSUSDH
- Nicole Shibata, CSUN
- David Walker, CO
- Jamie Zeffery, CSULA
- Digital Archives Working Group
    - Steve Kutay, CSUN, Chair
    - Nicole Shibata, CSUN, Vice Chair
    - Amalia Castaneda, CSUDH
    - Ariel Hahn, Cal Poly Pomona
    - Jill Strykowski, San Jose State
    - Adam Wallace, Fresno State
    - Jamie Zeffery, CSULA

Executive Summary

Pilot Background

From April through September of 2022, a pilot was conducted in order to identify development priorities for the shared digital archives repository of the CSU. Six campuses (CSUCI, CSUSM, CSUDH, CSULA, CSUN, and Sonoma State) voluntarily participated in the pilot assessment by contributing a combined 6,546 items and associated metadata records for ingest into an out-of-the-box (basic) DAMS installation of Samvera/Hyrax. Due to changes in campus personnel, one campus was unable to continue into the assessment phase. Assessors were asked to test applicable functionalities as outlined in CSU Libraries DAMS Report: Requirements for Implementing Shared Digital Library Services. A report form was created to capture assessors' data by asking them to review each functional requirement against a preset rubric by labeling each functionality using one of the following terms: 1) Not functional, 2) Does not meet requirements, 3) Meets requirements, and 4) Optimal – Advanced features. In addition, assessors were asked to supply text descriptions of the issues they identified as well as submit a screen captured image as necessary. For more on the construction of the pilot, see 'Methods', page 12.

Objectives

A previous evaluation of Samvera/Hyrax was published in 2022 entitled, *Assessment of Samvera/Hyrax for CSU Digital Collections* and was intended to gather published information and perform independent testing on the product and its associated technologies to determine how well the repository software would serve the needs of a shared CSU repository for digital archives. The report identified areas of concern, but identified reasonable solutions for all of the required functions and most of the preferred functions that were deemed "not functional" or "does not meet requirements". These solutions rely on currently available technologies outside the *basic* installation that integrate specifically with Hyrax, or could be programmed by third party developers.

This report differs from the previous in its focus and objective, which are: 1) the assessment is limited to the basic (out-of-the-box) installation of Hyrax version 2.9, and 2) this assessment serves to develop a priority list for development of the system to deliver the needs as presented by the functional requirements and the CSU overall.

Summary Analysis

There are numerous benefits for using Samvera as the DAMS for CSU Digital Archives. Among these are:

- Open source technology distributes development across a growing community of programmers.
- Modular construction means updated/advanced services can be swapped in rather than migrate everything to a new system.
- Built-in NDSA level 2 preservation support.
- With only one system to be supported for all participating campuses, overall maintenance and costs are reduced.

Pilot assessors were asked to lend their time toward identifying areas in need of attention. Some of their comments extend beyond the system itself to other related areas. Among the most important are:

- Further develop usage and analytics capabilities. Samvera comes with a dashboard and can produce usage and progress reports regarding items added. Assessors found the default capabilities need to be extended into more refined stats and reports. Specifics on this have not been articulated.
- Refine metadata to both be normalized across all participating campuses, while respecting insofar as possible, the unique needs of each campus.
- Refine default advanced search fields and navigation to be more user friendly and more appropriate to searching archival materials.
- Refine roles and workflows to fit across varying needs of campuses across the CSU, which will require both feedback and teams to build these out.
- Compound objects metadata are currently difficult to navigate. Child records are underdeveloped and should be present alongside the master (or mother) record.
- Provide a platform for constructing exhibits with heavy contextual descriptions. Existing plugins that enable this are currently available, but not installed as part of the basic installation and therefore not assessed in the pilot.
- Current document viewers and players work generally well, and conform to IIIF (International Image Interoperability Framework) standards. However, more advanced IIIF compliant viewers are available that can be easily integrated with the system.
- Book and newspaper readers (flip readers) are not native to the basic system installation but can be added later to improve the user experience of viewing the materials.

Data are summarized below for your convenience.

Summary Data: Backend (Administration) Results

Compliance ratings given next to each functional section with comments summarized on individual functionalities. For more details, see "Pilot results – Backend", page 16.

(R) = Required functionality; (P) = Preferred functionality

User Management – Meets requirements

- Further testing for performance refinement recommended.

Content Types – Does not meet requirements

- File formats (R): Problems uploading audio files (not video files).
- File swapping (P): When swapping files, titles and technical metadata do not immediately update or do not update at all.

Object Handling – Does not meet requirements

- Derivatives (P): TIF file is rendering as the access file.
- Permanent URL (P): No permanent URL function. Must use ARK or other system.
- Embargo (P): Inconsistent performance on embargo dates.

Modeling (Data) – Meets Requirements

Metadata – Does not meet requirements

- Common Schemas (R): Limited. Only DC enabled. Must survey users if other schemas are required.
- MODS (P): No support. Could be developed on demonstrated need.
- Controlled Vocabularies (R): None. Requires Questioning Authority gem installation.
- Linked Data(P): None. Requires Ruby gem installation.
- Remediation/Editing (R): Confusing/competing date fields. More testing on date performance for searches.
- Find & Replace (R): None. Requires Bulkrax plugin installation.
- General:
    - Title, Identifier, and Description labels not displayed on metadata records. Some field labels not intuitive.
    - Performance of date search unclear. Not clear which date fields (created, published, issued, etc.), are being extracted for date facet. Establish system for date ranges.

Ingest – Does not meet requirements

- Bulk Ingest, simple (R): Not tested. Requires Bulkrax installation.
- Bulk ingest, compound (P): Not tested. Requires Bulkrax installation.
- Add object Interface (P): Individual item adding is enabled.
- Workflow moderation (R): Frustration with requirement that some fields MUST be completed before being able to save incomplete work. Cannot save without an associated file.

Quality control (P): Meets requirements

- Must refine user permission to allow for greater project manager access.

Exports – Not tested in pilot

For more on these features, see previous report for evaluation of this area.

- Full metadata export (R): Requires Bulkrax.
- Selected metadata export (R): Requires Bulkrax.
- Multi format metadata extraction: Only CSV available by default. More development required.
- Technical metadata extraction (P): Enabled via FITS toolset on ingest.
- Bulk file export (P): Requires Bulkrax.

Back-up/Sustainability – Not tested in pilot

Note: these features run in the background or are external to the application.

- Redundancy (R): Uses AWS (Amazon Web Services).
- Recoverability (R): Failures confined within the system stack. AWS backups. Fixity checks on ingest.
- Fault tolerance (P): Distributed servers enable all unaffected services to continue. For more, see Results section of this report.
- Network storage (R): Uses AWS.

External System integration – Not tested in pilot

These features are administered at the CO level

- OAI-PMH (R): Currently Implemented, but will require additional configurations.
- ULMS Harvest (P): A customization solution already created for ScholarWorks.
- API Support (R): Enabled.

Reports – Not Tested in pilot

Outside of scope of the initial assessment, but may be investigated later for further development.

- Analytics (R): Basic usage statistics available through dashboard.
- Technical/Administration: Basic reports for items and collections are functional.

Preservation Management – Not tested in pilot

These features run in the background and/or limited to upper level administrators of the system.

- File redundancy (P): Supported.
- XML/METS (R): Not supported, but could be developed if a demonstrated need exists.
- Fixity (P): Supported.
- Format obsolescence (P): Not automated, but can accommodate file swaps from Fedora reports.
- Born-digital ingest: Supported.

Summary Data: Frontend (End-user) Results

Compliance ratings given next to each functional section with comments summarized on individual functionalities. For more details, see "Pilot results – Frontend".

Discovery – Does not meet requirements

- Advanced search (R): Link to advanced search page need to be prominent in the home page. Default search fields must be changed to meet needs of archival document searching, not books. Some fields are not high priority for advanced searching.
- Title, keyword, author search (R): No comments.
- Customizable filters (R): Customization is supported. Date facet pulling from date of publication, not date of creation as would be expected for archives.
- General: Sort by date (created) not enabled.

Delivery Interface – Meets requirements

- Zooming, Downloading, Printing (R): No print button for PDF, only download option, which is required to view the PDF in browser. Prefers different options for viewing compound objects.
- Advanced analysis (P): No comments.
- Display sizing, rewind, fast forward (R): No comments.

Web pages – Split (Meets/does not requirements; Not functional)

- Displaying Lists (R): No comments.
- Textual descriptions (R): No comments.

- Campus branding (P): No comments.
- Web standards (R): No comments.
- Accessibility (R): No comments.
- Image rendering/Object embedding (R): Permanent URLs needed for object embedding on PDF.
- Flip reader (P): No flip readers, but can be added.

Exhibits (P) – Split (Not functional; Does not meet requirements)

- Tile and slideshow views are beneficial for exhibits, but requires exhibit plugin Spotlight for contextual descriptions.

Downloads – Meets requirements

- Multiple Resolution Downloads (images) (P): No comments.
- Printing (P): No printing function, however download function will allow users to print from computer.

User tags/Annotations – Not functional

- Tag items (P): No ability to tag items in list and other views.
- Annotate (P): No comments.

Social media support – Meets requirements

- No comments.

Methods

Pilot structure

During the spring through summer of 2022, six CSU campuses agreed to participate in the testing of an out-of-the box installation of the Samvera-Hyrax digital asset management system to construct a set of development priorities to quickly bring the system in line with the functional requirements assembled by the Digital Archives Task Force.

The following schedule outlines the phases of the pilot:

1. April – Campus invitations to participate announced on Digrepo listserv. Project kick off meeting to outline the objectives and establish instructions for executing the pilot. Testers review the assessment rubric and additional tools for reporting their findings.
2. May – Configure and review baseline metadata application profile to comply with harvesting requirements of the California Digital Library and Digital Public Library of America.
3. June – Begin testing. Pilot participants were asked to assemble test collections from existing collections by selecting a variety of file formats in order to test all aspects relating to the ingest, delivery, and handling of text, image, sound, and moving image files.
4. July – Check in to review progress and answer questions.
5. August – Check in to review progress and evaluate further needs.
6. September – Testing complete.

Rubric

In preparation for a CSU shared digital repository, the Digital Archives Working Group created a rubric to evaluate compliance with the 2019 DAMS functional requirements report. Given differences in functional complexity of any single feature in a DAMS, the tool deploys three strategies within its rubric. A binary scale applies to services that are simple, that is, they are either functional or not functional, such as authentication. A trinary scale applies to features that are more nuanced and could potentially be developed beyond our minimum requirement, thus a rating of 'Optimal - Advanced features' could be assigned. A third, granular strategy, is applied to complex features, such as advanced search capabilities, that could be functional, but still not meet our minimum requirements. See figure 1.

	Scoring (points applied)			
Strategy	0	1	2	3
1 (binary)	Not functional	X	Meets requirement	X
2 (Trinary)	Not functional	X	Meets requirement	Optimal (Advanced Features)
3 (Granular)	Not functional	Does not meet requirement	Meets requirement	Optimal (Advanced Features)

Figure 1: Scoring rubric features multiple strategies depending on the complexity of the function assessed.

Digital Repository Assessment

The rubric is accompanied by the Digital Repository Assessment Tool created by the Digital Archives Working Group specifically for facilitating early assessments of digital asset management systems (DAMS) under consideration or development for serving the needs of CSU Libraries. This assessment tool

informed the structure of a problem report form (see Pilot Report form), which was used to gather additional qualitative data from the testers.

Collection base URLs used in this assessment		Discovery			
		Advanced search (R)	Title, keyword, author, se	Customizable filters (R)	NOTES
	Rubric Strategy deployed	3	3	3	
http://ec2-34-210-218-97.us-west-2.con	**Assessment score**	Meets requirements	Meets requirements	Meets requirements	
	Numerical Score	2	2	2	
	Points possible	3	3	3	
	Difference	**1**	**1**	**1**	**Add notes here**

Figure 2: Capture of the Digital Repository Assessment Tool. Example of assessed category with subcategories. Selection of assessment score auto-populates a numerical score.

The pilot report form was the primary means for recording the pilot assessment, which represents the previously articulated needs from the Digital Archives Working group report, *CSU Libraries DAMS Report: Requirements for Implementing Shared Digital Library Services.* Like the assessment tool, the pilot report form defines functional requirements, which are divided into functional areas for which the testers rated using the following labels: 1) Not functional; 2) Does not meet requirements; 3) Meets requirements, or; 4) Optimal - Advanced features.

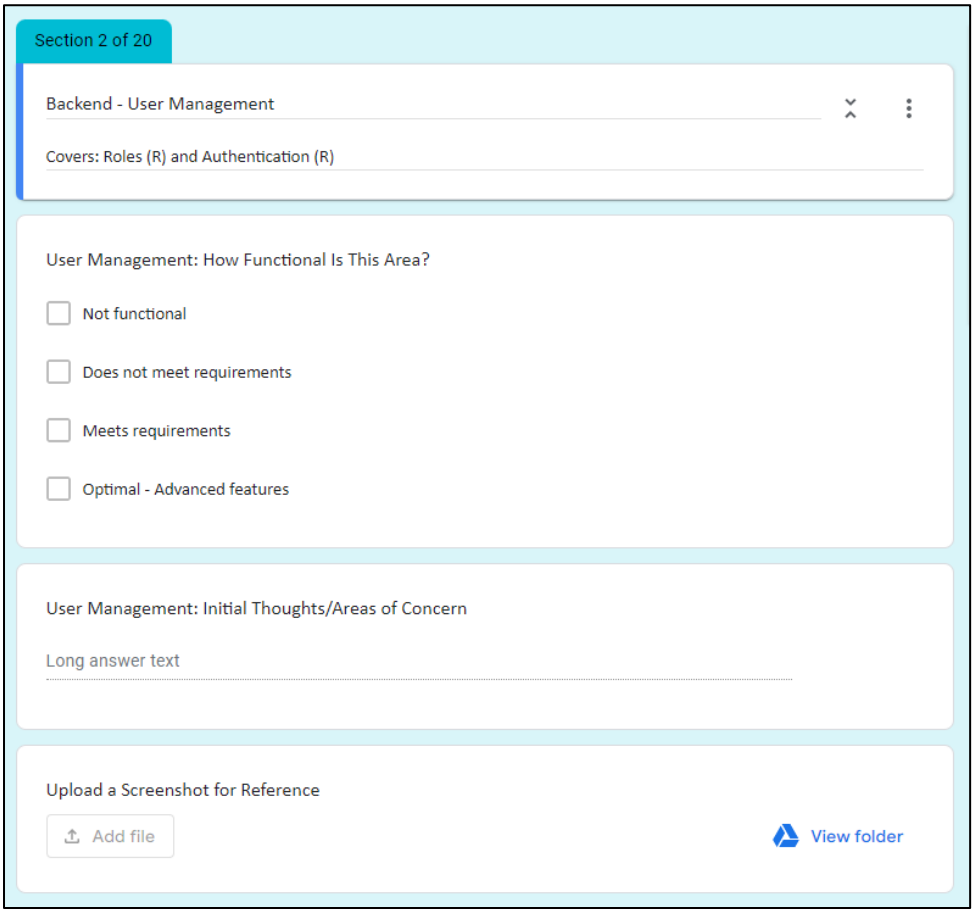

Figure 3: Screencap of the report form for the User Management feature set.

For more on the evaluation of Samvera/Hyrax for CSU digital archives, see the related preliminary report, *Assessment of Samvera/Hyrax for CSU Digital Collections*.

Referenced Resources

Pilot participants were directed to the following online resources to assist with learning about Samvera/Hyrax in order to conduct this assessment:

Backend

- Basic (out-of-the-box) Hyrax v3.0 installation at CSU San Marcos - https://digitools.csusm.edu/ Does not include full technology stack.
- Samvera Knowledge Base - https://samvera.github.io/index.html
- About Hyrax (for Managers) FAQ - https://hyrax.samvera.org/about/managers/
- Fedora - https://duraspace.org/fedora/

Frontend

- Basic (out-of-the-box) Hyrax v3.0 installation at CSU San Marcos - https://digitools.csusm.edu/ Does not include Universal Viewer.
- Spotlight - https://demo.projectblacklight.org/
- CSU ScholarWorks - https://scholarworks.calstate.edu/
- Spotlight at Stanford - https://exhibits.stanford.edu/
- Universal Viewer - https://universalviewer.io/

Limitations

Limitations to this evaluation of Samvera/Hyrax for CSU Library digital collections are as follows:

Participation

The assessment was open to all campuses; however, this evaluation was conducted by volunteer participants who represent only six of the twenty-three CSU campuses.

Resources

Time and commitment to the evaluation were limited, given the continued professional responsibilities of the evaluators.

Independence

To preserve integrity, the evaluation was conducted independently of direct input from developers and users of various open source components of the Samvera technology stack and associated alternative applications, plug-ins or "gems". As such, gaps in this analysis exist that could otherwise be filled with direct inquiries to developers and users of Samvera and associated technologies in order to clarify ambiguous findings identified within this report.

Objectivity

Despite the application of a rubric and assessment tool to guide this evaluation, the deployment of individual rubric strategies and the subsequent scores selected by the evaluators can be subjective. Relative experiences of any evaluator can lead to somewhat different results.

System Costs

Typically, Library and Information System service costs are negotiated. Despite the importance of costs to participating CSU campuses, the financial impact of deploying Samvera as a shared DAMS for the California State University System was not a part of this evaluation. However, the Digital Repository Committee, in conjunction with the associated working groups have flagged this as a fundamental priority requiring a separate investigation.

Pilot Results - Backend

Testers were asked to review each functional requirement within the designated functional categories that were available to them and comment on areas in need of improvement. Administrative functions are listed here for reference, but not tested as part of this assessment.

(R) indicates a required feature, (P) indicates a preferred feature

User Management

Roles (R)

Definition: User roles should include: 1) Super-User (or administrative equivalent) that has complete configurable access; 2) Project Administrator to provide the highest levels of database maintenance and assign users with subsequent permissions; 3) Users with levels of regulated access to specific features and collections in the system, and; 4) Public User accounts to enable personal tagging/organization of online items.

Authentication (R)

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

Compliance rating: Meets Requirements.

- Further testing required.

Comments

- It is not clear what features are available at each user level. It seems that my setting is "Project Administrator" (My profile information doesn't indicate it and I wonder how I can find out which user level I am…?). I am not able to edit the records uploaded by the Super-User but it is inconvenient and does not make sense to me. I was able to test "Project Administrator" and "Public User" but need to test "Users." Can I have the User level access to see how it works? Authentication options are available but need to verify if they are actually functioning.
- Roles need to be verified by using different level accesses. Authentication: "Private" restricted access works well. Only staff can view and access the restricted files. Need to verify other settings, such as campus and CSU.

Content Types

Format Agnostic (R)

Definition: Both born-digital and digitized content will be migrated to digital collections environments. Therefore, the DAMS must be format-agnostic to ensure all file types can be managed as needed.

File Swapping (P)

Definition: In the event an ingested file must be changed, the system should accommodate the swapping of new versions of files.

Streaming (R)

Definition: System should effectively serve streaming media content of audio and video formats.

Compliance rating: Split (Does not meet requirements/Meets requirements)

- Further audio testing and required.

Comments

- When I replaced a TIFF file with a JPEG file, the file was replaced and file details were updated but the file name remains intact in the Title. I was not able to upload an audio file. I do not know if the file size was too large or the system does not accept an audio file.
- File swapping: Files can be swapped with new versions and also older versions can be restored. However, embedded metadata (file details) is not updated accordingly. I thought the details were updated before…? Streaming: A mp4 file that I uploaded works well. I was not able to upload mp3 for testing. I have been getting an error.

Object Handling

Derivatives (P)

- Definition: The processing of preservation quality formats as access derivatives, such as TIFFs to JPGs, will provide an optional workflow for digital preservation of analog materials as well as valuable automation.

Object Rights Management (P)

- Definition: Object level rights management and branding should be enabled through the automatic processing of watermarks and/or embedded banners at the time of ingest. (P)

OCR/Full text (R)

- Definition: The system should provide automated OCR at ingest for full-text searching of applicable textual materials with typeface.

Permanent URL (P)

- Definition: Each item record is associated with a fixed and permanent universal resource locator (URL) to ensure external links to an item record are not broken.

Embargo (P)

- Definition: Enables the ingest of items that are subsequently suppressed from viewing for a predetermined period of time in order to accommodate issues pertaining to copyright, or permission to publish.

- Average score: 2/2, Meets preferred feature
- Can manage embargoes as well as leases.

Compliance rating: Does not meet requirements

- Further development required.
- Inconsistent derivative production.
- It looks like the entire TIF file is rendering as the access file. Is it possible to set it so that the access file is a (automatically generated) jpeg?
- No permanent URL.
- No out-of-the box OCR.
- Fix embargo.

Comments

- Derivatives: I would like to have this function available but it seems not functional. Object Rights Management: Some aggregators (Calisphere) do not recommend watermarks or brands. OCR/Full text: I do not think that it is functional...? Permanent URL: Endnote generates a permanent URL, but it doesn't work. A "share" icon is available for images and videos but not for pdfs. I would like to create a permanent URL pointing to search results. Would it be available?
- Derivatives: Uploaded TIFFs were not converted to JPEGs. It would be useful if the Hyrax system allows users to select whether TIFF is converted to JPEG or uploaded as TIFF. It would be ideal if TIFF is saved for staff use while JPEG is used for the public. OCR/Full text (R): Uploaded PDFs (no-OCR) were not OCRed automatically during the ingest. Embargo (P): Not functioning. I set a date, 2022-09-08, it is not available on September 9, 2022.
- Downloadable TIFF—can JPEG2000 be generated from TIFFs? Or do we need to create JPEGs for downloading?
- Tiffs - We cannot allow download of many of our tiffs, due to copyright restrictions. Can tiffs be displayed, but jpegs be generated for download?

Modeling

Data modeling (R)

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

Compliance rating: Meets requirements

- Further development requested on a more intuitive display of compound object (mother-child) metadata records.

Comments

- The compound object structure is a little questionable. It is hard to tell if the item is a compound object or a multiple-page structure. The problem I face is that the child record can be viewed from the parent record, but the parent record cannot be viewed from the child record once you

click on the child record. That is, the relationship between the parent and child record is not retained or displayed well.

Metadata

Common Schemas (R)

- Definition: Descriptive schemas Dublin Core, VRA Core 3.0, PBCore are commonly used for non-bibliographic materials.
- Supports simple Dublin Core by default, but can integrate others if a campus can provide a case need and maintenance support.

MODS (P)

- Definition: Bibliographic descriptive schemas such as MODS will add to the extensibility of the platform. Incorporate nested data.
- NOTE: Can be configured for MODS, however doing so requires continual maintenance for future updates.

Controlled Vocabularies (R)

- Definition: As a best practice of resource description, the DAMS system must enable the management of controlled vocabularies derived from both standard thesauri and local controlled vocabularies.
- NOTE: Uses the Questioning Authority gem that comes configured for most common vocabularies. Has the ability to create custom vocabularies, query standard vocabularies and utilize linked data access. However, methods are not intuitive and will require implementation of tools or workflows to assist most users.

.

Linked Data (P)

- Definition: Linked data implementation that provides direct access to standard vocabularies will promote consistent descriptive practice between items and collections.
- NOTE: Uses Ruby gem LDP as linked data platform for connecting with Fedora. Full features in development.

Remediation/editing (R)

- Definition: The platform should offer object level editing capabilities to facilitate metadata updates to ingested materials.
- NOTE: Metadata can be remediated at the object and collections levels.

Find & Replace (R)

- Definition: The platform should offer global metadata editing capabilities within specific fields, such as searching and replacing existing values within a single field, as well as blanket changes to entire field values, such as rights statements.
- NOTE: Uses Bulkrax as a metadata editor, scheduler, and reviewer. Contains an admin interface. ScholarWorks has also made implementations. Bulkrax NOT installed at the time of this test.

Compliance rating: Does not meet requirements/Meets requirements (split)

Comments

- Title, Identifier, and Description labels are not displayed on the public interface. There are labels for “Alternative title” and “Other identifiers” but there are no labels for the “main” title and identifier.
- It seems that the date facet is extracting data from “Date of publication” which is available only after 1940. It should extract data from “Date created” or the date of publication should cover years earlier than 1940. If you do not enter “Date of publication,” it seems that the date facet extracts data from “date modified,” that is, 2022.
- Some labels are too technical and not user friendly. Using specific words, such as, “Finding Aid” for “Is version of,” “Related Collections” for “Relation,” would be useful.
- Date fields with date ranges migrated from CONTENTdm are displaying with every single year within that range listed. I've attached a screen shot. The nonmachine-readable field reads correctly as "1915-1930" but I'm wondering if Hyrax is actually indexing all-inclusive dates here.
- Find & Replace/Fill would be ideal but is it available for testing? Collection/Sub collection level metadata: It seems that the Hyrax allows users to set another set of metadata at the collection and sub collection levels respectively…?
- Would linked open data be available?

Ingest

Bulk ingest, simple (R)

- Definition: The system should provide batch (or bulk) ingest of multiple items of simple objects of files and metadata.
- NOTE: Uses Bulkrax as importer, scheduler, and reviewer. Contains an admin interface. ScholarWorks has also made implementations. Bulkrax NOT installed at the time of this test.

Bulk ingest, complex/compound (P)

- Definition: The system should provide batch (or bulk) ingest of multiple items of compound objects of component files and metadata.
- NOTE: Multiple files can be uploaded as separate works, and for each separate work multiple component files can be uploaded as associated with an individual work to form a compound object. However, this, by definition, is not a bulk ingest of compound objects. Bulk compound ingest could be flagged for development if a need case is demonstrated.

Add object interface (P)

- Definition: The system should provide a user interface for adding individual objects (simple or compound) directly into the system.
- NOTE: Our testing indicates an effective add object interface and workflow.

Workflow moderation (R)

- Definition: Can save and revisit projects before completing ingest or updates. (R)
- NOTE: Ingest must complete, however they can remain unpublished and edited any number of times before and after publishing. You can change the structure of a work from simple to compound by associating new files with the work after ingest.

Testers did not test bulk ingest since this was a Super User function for which only CO staff was assigned at the time of this pilot. Testers DID add individual objects manually and commented accordingly.

Comments

- Unless all three steps (Describe your work; Add files; and Check deposit agreement) are filled, the system does not allow to save incomplete work. This is inconvenient.
- Unless all system requirements, "Describe your work," "Add files," and "Check deposit agreement," are filled, the item would not be saved. The system does not allow users to save incomplete items, such as metadata only. It is a little inconvenient.
- “Create batch of works” does not work.
- Save incomplete metadata record/Save metadata record without file - It would be great to be able to save an incomplete metadata record so it could be finished at a later date. Same with saving a metadata record without a file.

Quality Control (P)

- Definition: The system should provide quality control functions to administrators for approving, removing, or updating submitted objects within a queue prior to the final stage of ingest, or the building of a collection’s index.
- NOTE: Uses mediated deposit to provide approval process. More complicated workflows are enabled via programming.

Compliance rating: Meets Requirements

Comments

- As I noted earlier (User Management), it is inconvenient if “Project Administrator” can’t edit the records uploaded by “Super-User.” I would like to test the User level access as well.

Exports

Full metadata export (R)

- Definition: A collection’s full metadata should be exportable for external remediation, preservation, or as downloadable files for public use such as research and machine learning operations.
- NOTE: Uses Bulkrax as exporter. Contains an administrative interface. ScholarWorks has also made implementations from which to model. Bulkrax NOT installed at the time of this test.

Selected metadata export (R)

- Definition: Results from a search of a collection’s contents should be exportable for external remediation, or entry into other collections within the system.
- NOTE: Uses Bulkrax as exporter. Can pull selected metadata for remediation and import back into the system. ScholarWorks has also made implementations from which to model. Bulkrax NOT installed at the time of this test.

Multi format metadata extraction (R)

- Definition: Exports should be offered in multiple formats such as delimited text, XML schemas, and METS to provide flexibility when working with databases, or for long-term preservation.
- NOTE: Uses Bulkrax. Supports export of common CSV format by default as of 2020. More format support is prioritized through Notch 8 developers. However, progress has not been confirmed at this time. Bulkrax NOT installed at the time of this test.

Technical metadata extraction (P)

- Definition: System exports should include the extraction of technical metadata of items within the database.
- Uses File Information Tool Set (FITS) upon file ingest to extract technical metadata.

Bulk file export

- Definition: System enables the export of files at the collection level, campus level, or results from a search.
- Uses Bulkrax as exporter. Can export metadata from importer, collection, or work type. Bulkrax NOT installed at the time of this test.

No rating: Not tested for this assessment, but will be beta tested once Bulkrax is applied.

Back-up/Sustainability

Redundancy (R)

- Definition: Redundancy of files and metadata to protect against catastrophic loss. (R)
- NOTE: Ingest copies created in Amazon Glacier. Another copy via AWS. Campuses may wish to maintain additional copies.

Recoverability (R)

- Definition: Databases must be fully recoverable in the event of catastrophic failure.
- The system as a technology stack has an advantage of limiting failures to any one specific technology (functional area) rather than rendering the entire system nonfunctional (see Fault tolerance). In the event of a catastrophic failure Fedora supports fixity checks (SHA-1, SHA-256, MD5) for single files and metadata and will compare against existing fixity digests. In addition, Glacier provides back-ups and fixity checks on ingest.

Fault Tolerance

- Definition: System should be fault tolerant to continue operation during erroneous or compromised performance. (P)
- Components exist on different servers which maintains the system in use when one process goes down. Only the service is impacted. Uses separate file server.

Network storage

- Definition: Can Connect to external storage environments within a network such as local drives, cloud/web services, etc. typically to complete ingest or swap files. (P)

- Can connect to cloud services for ingest of existing files. Unclear whether it can store preservation files or derivatives to external storage such as AWS.

No rating: Administrative only. Not tested as part of this assessment.

External System Integration (Interoperability)

OAI-PMH (R)

- Definition: The system must enable the crawling of a collection's metadata and URLs as desired via OAI-PMH to be used within external discovery systems such as Calisphere and the Digital Public Library of America.
- NOTE: Currently Implemented. See Blacklight OAI-PMH. Partial with Fedora bag support. Does not fully support harvesting out of the box, but can be configured with other applications. Will provide imports via OAI-MPH.

ULMS harvest (P)

- Definition: Records within the system should be harvestable by the CSU Unified Library Management System (Ex Libris Alma) via an external API or another protocol.
- Blacklight OAI Provider plugin can be implemented. A customization solution already provided for ScholarWorks.

API Support (R)

- Definition: Support available to campus personnel that create customizations. (R)
- Generally, CO personnel will provide API support. Will consider allowing other campus personnel contribute support where applicable. Uses customizable HTTP APIs to communicate with Fedora (storage) and Solr (index).

No rating. Administrative only. Not tested as part of this assessment.

Reports

Analytics (R)

- Definition: Collection-level and item-level usage statistics and analytic reports will provide valuable data to enable administrators to strategize development and promotion of their collections.
- NOTE: Multiple stats available via dashboard. Leverages Google analytics. Limited, but new module coming soon, according to DLS Director.

Technical/Administration (R)

- Definition: Reports such as items added, file types, objects per collection, users, etc.
- Previous assessment indicates an acceptable variety of reports to cover users, collections and objects.

No rating: Not tested as part of this assessment, but will be available to future beta tests.

Preservation Management

File Redundancy (P)

- Definition: Redundant file storage of preservation files

- Average score: 2/2, Meets preferred feature
- NOTE: Ingest copies of preservation files in Amazon Glacier.

XML/METS (R)

- Definition: Structured metadata exports in XML/METS (see Exports)
- NOTE: Nonexistent. Could develop if a case need is demonstrated.

Fixity (P)

- Definition: Integrity checksums (e.g. MD5, SHA256)
- NOTE: Fedora supports fixity checks (SHA-1, SHA-256, MD5) for single files. Will compare against existing fixity digests. Glacier provides fixity checks on ingest.

Validation (P)

- Definition: Format validations (e.g. JHOVE)
- NOTE: Uses FITS through Hyrax File Characterization. JHove and DROID included in FITS. Extracts technical metadata.

Format Obsolescence (P)

- Definition: Automated reformatting of obsolescent formats (P)
- NOTE: Obsolescence monitoring or automated format migrations are not expected. Fedora can be queried for file formats in order to inform preservation management decisions, which is accomplished via file swaps.

Born-digital ingest (P)

- Definition: Ingest of born-digital materials for preservation
- NOTE: System is format agnostic and capable of ingesting all file types (including files/data created within a born-digital domain). However, the system does not serve as a standalone preservation management system. Preservation priorities and workflows for migrating and managing preservation files and metadata (including from fixed storage systems) are the responsibility of each campus. Born-digital preservation management guidelines and resources have been created by the Digital Archives Working Group for voluntary use by CSU campuses.

No rating. Administrative only. Not tested as part of this assessment.

Results – Samvera/Hyrax Frontend

Testers were asked to review each functional requirement within the designated functional categories that were available to them and comment on areas needing improvement. Administrative functions are listed here for reference, but not tested as part of this assessment.

(R) indicates a required feature, (P) indicates a preferred feature

General note: As a technology stack, Hyrax uses Blacklight and Spotlight as its web interfaces. They are able to implement a variety of IIIF viewers and audiovisual media playback engines to enable user searching and delivery of materials within a digital asset management system. Only out-of-the box configurations were tested here, though others can be added/swapped to improve performance. The assessment of specific functional requirements for the DAMS frontend are below.

Discovery

Advanced search (R)

- Definition: Search engine should provide advanced search (multiple Boolean and specified field search bars). Search across specific fields.
- NOTE: Hyrax v3 out of the box, does not offer multiple search bars using Boolean operators and variable field assignments. However, the CO is planning to build this feature based on the ScholarArchive@OSU.

Title, keyword, author, search (R)

- Definition: Search engine should provide, specifically, keyword, author and title search capabilities.
- NOTE: Search box offers field specific searching for 'All items', 'My Works', and 'My Collections'. Other specific values must be added.

Customizable filters (R)

- Definition: Discovery layer should provide customizable filters and facets. (R)
- NOTE: Uses Blacklight, which truncates long facets, but enables popups for complete lists. Contains graphic display of time distribution of documents. Out of the Box Hyrax V3 utilizes facets.

Compliance rating: Does not meet requirements

Comments

- Advanced search looks hidden. The link to the advanced search page needs to be displayed at the first landing page.
- Selected fields for advanced search are too book-oriented.
- I do not know if “Publisher” is necessary for archives–items are often unpublished.
- I do not know if the filters and faceting can be customizable—if they are, that would be great.
- It seems that the date facet is extracting data from “Date of publication” which is available only after 1940. It should extract data from “Date created” or the date of publication should cover years before 1940. If you do not enter “Date of publication,” the date facet extracts data from “date modified,” that is, 2022.

- Sorting options: “date uploaded” and “date modified” are not as useful for users; title and date sorting options are more useful.
- Customizable filters/faceting: Unable to verify it. It seems that the date facet was extracting data from “Date of publication” which was available only after 1940. It should extract data from “Date created” or the date of publication should cover earlier years. If you do not enter “Date of publication,” the date facet extracted data from “date modified,” that is, 2022. These date fields are no longer visible now on the backend.
- Sorting options: “date uploaded” and “date modified” are not as useful for users; title and date sorting options are more useful.

Delivery Interface

Zooming, Downloading, Printing (R)

- Definition: PDF viewers should be enabled and effective regarding zooming, downloading, and printing.
- NOTE: No direct-to-printer icon, but offers download from which to print. One can be implemented if desired.

Advanced Analysis (P)

- Definition: IIIF and other embedded viewers will facilitate advanced analysis of image files.
- NOTE: Contains gem RIIIF to install the IIIF universal viewer (https://universalviewer.io/) More advanced viewers, such as Mirador (https://projectmirador.org/) and OpenSeadragon (https://openseadragon.github.io/) can be installed by consulting the work of other institutions (https://openseadragon.github.io/examples/in-the-wild/).

Display sizing, rewind, fast forward (R)

- Definition: Audiovisual media players should enable basic functions such as display sizing, rewind and fast forward.
- NOTE: Uses IIIF Universal Viewer for image display and audiovisual playback. Other IIIF viewers can be implemented.

Compliance rating: Meets requirements

Comments

- No PDF viewer not intuitive. Requires that “Download” button be clicked and viewed in browser.
- A child record of a compound object needs better navigation to access its parent record.

Web Pages

Displaying Lists (R)

- Definition: The system should provide the means for displaying lists of collections (e.g. splash pages) as well as the means for displaying these according to each participating campus. This could be a native function of the software and/or a customizable web-based option.
- NOTE: Can display collections across all campuses. Since collections can be nested and items can exist in multiple collections, exhibits can use items from multiple campuses. In addition,

metadata will identify contributing campuses therefore, the built-in facet system can be used to filter campus collections.

Textual Descriptions (R)

- Definition: The system should provide the means for context through textual descriptions (e.g. landing pages) of the collections in order to give users critical context associated with the materials contained therein. This could be a native function of the software and/or a custom web-based option, such as a WYSIWYG editor.
- NOTE: Collection descriptions are supported. Advanced descriptions are enabled by the plugin, Spotlight. Furthermore, collections nested within a collection are supported. This has the potential for shared campus exhibits of archival collections of different provenances across the CSU.

Campus Branding (P)

- Definition: The system should provide the means for campus branding of the collections through graphics, wordmarks, and color assignments
- NOTE: Basic branding is supported. Colors and logos are available in Hyrax and enabled by the CO.

Web Standards (R)

- Definition: The user interface must be responsive according to web standards.
- Basic branding is supported. Colors and logos are available in Hyrax and enabled by the CO. Fedora complied with modern web standards.

Accessibility (R)

- Definition: The system must be compliant with the Americans with Disabilities Act (ADA) to ensure accessibility, or is committed to reaching full compliance.
- NOTE: Blacklight accessibility is likely mixed. There are no known (public) comprehensive tests of ADA compliance. However, the Blacklight plugin, GeoBlacklight, claims to fulfill WCAG 2.0, Section 508, and other accessible standards. Beyond this, there is at least one claim in the library community that Blacklight has performed reasonably well among accessibility tests, with some development still required. Accessibility developments have been made for ScholarWorks. Hyrax contains task forces to address accessibility issues. Will need to consider accessibility development locally based on needs for a shared digital collections DAMS.

Image Rendering/Object Embedding (R)

- Definition: With each item created, the system should be able to associate that item with a direct link to the fully rendered object file that can then be reused in third party tools (e.g. exhibit platforms, webpages) for image rendering and object embedding using a stable URL.
- NOTE: Service not currently available, but can be customized on demonstrated case need.

Flip Reader - Book/Newspaper Readers) (P)

- Definition: Special object display such as side-by-side page flip view, or news article highlight/selection.
- NOTE: Uses Newspaper Works, a rails gem to serve newspaper content. Contains both back and front-end feature sets.

Compliance rating: Split (Meets requirements. Does not met requirements, Not functional)

Comments

- As noted earlier (Object handling), a "share" icon generates a permanent URL and is available for images and videos but not for pdfs. Endnote also generates a permanent URL but it is not functional.
- No Flip reader. Open source flip readers can be implemented. Waiting on implementation of Rails gem, Newspaper_Works.

Exhibits (P)

- Definition: As an identified priority of the survey respondents, it is requested that the system provide the means for curating exhibits online. One complicating factor is that no mutually agreed upon definition of what constitutes an exhibit online was achieved. Some digital asset management systems offer the means to create web pages, however limited, toward an exhibit-like function. In the absence of a dedicated exhibit function, the system should utilize plug-ins, extensions, or integrated applications designed for this purpose. (P)
- NOTE: Uses Spotlight to create feature rich websites for curating digital collections materials. As an open source technology, newly developed widgets, etc. can be implemented, making the platform extensible moving forward.

Compliance rating: Split (Not functional; Does not meet requirements)

Comments

- System provides tile and slideshow views of items that are useful for exhibits. However, waiting on Spotlight implementation for full exhibit functionality.

Downloads

Resolution Downloads (P)

- Definition: The system should provide full and partial resolution downloads of the materials within the database to enable flexible use in research and course learning objectives.
- NOTE: The system does not offer the means to select and download varying resolutions of images. Downloads are enabled at full resolution. Customization of this service could be requested on demonstrated need.

Printing (P)

- Definition: The system should enable printing of reasonably of file types within the database.
- NOTE: Only downloads are offered, but as an alternative, the item can be printed upon download. Given this knowledge, direct printing is a convenience, and therefore this feature has been downgraded to "Preferred" (P).

Compliance rating: Meets requirements

Comments

- There is an icon for downloading but it is not downloaded as a file.
- There are no icons or options for printing, but can be printed via download.
- No direct download, but moves file to new tab to perform "Save file as".

User Tags/Annotations

Tag items (P)

- Definition: User functionality is greatly improved by the ability for the public to tag individual items for future use. (P)
- NOTE: Items can be bookmarked in the gallery view.

Annotate (P)

- Definition: Users and custodians of archival materials benefit significantly from the ability of users to annotate. (i.e. provide public descriptions) of the database items therein. (P)
- NOTE: No ability to annotate objects within the system. However, some IIF viewers can accommodate certain annotations. Also, the system can provide links to web forms in order to submit public descriptions to specific personnel.

Compliance rating: Not functional

Comments

- Should be able to tag/bookmark items when using the gallery view, but not seeing this.

Social Media Support (P)

- Definition: The system should enable sharing through social media by building in links to popular social media sites accessible at the item level through the use of stable URLs and embed codes, or in the least, provide easy access to stable URLs and embed codes. (P)
- NOTE: Facebook, Twitter and Tumbler are enabled out of the box at the item level, new social media links can be customized.

Compliance rating: Meets requirements

No comments

Findings and Conclusion

See Executive Summary, page 7.

Send questions or comments to stephen.kutay@csun.edu.