

Next-Generation Technical Services (NGTS)

Digital Asset Management System (DAMS) Requirements

Final Report

July 20, 2012

Power of Three (POT) #1, Lightning Team #1A

Kathleen Cameron, UC San Francisco (Co-chair)

Declan Fleming, UC San Diego (Co-chair)

Giulia Hill, UC Berkeley

Adrian Turner, California Digital Library

Mary Wepler-Selear, UC Merced

Audra Eagle Yun, UC Irvine

Contents

Executive Summary	3
Methodology	3
Interview Participants	4
Working Definitions	4
Content Contribution Requirements	5
Aggregation Requirements	7
Discovery and Display Requirements	8
Addendum A: Interview Questions	10

Critical Requirements for a UC Systemwide Infrastructure for Digital Collections
UC Libraries Next Generation Technical Services POT 1, Lightning Team 1A
July 20, 2012

Executive Summary

The Power of Three 1 (POT1) group was charged with identifying the critical requirements for a systemwide DAMS that meets the needs of those UC campus libraries that do not yet have one that would include the following:

- Determine access control needs.
- Determine branding needs.
- Determine which formats need to be supported by DAMS with discovery and display
- Determine if the effort should go toward a single, hosted DAMS with discovery and display, many DAMS with discovery and display that are installed at individual campuses, or some hybrid.

In May 2012, Lightning Team 1A (LT1A) was assembled and determined that eight out of ten campuses are very interested in participating in a central, systemwide solution. The majority of campuses felt that the solution should:

- Treat content files as agnostic
- Be able to display simple and complex object types (one digital object)
- Allow for easy access to descriptive metadata
- Allow for different editing levels for different roles by types of staff
- Authenticate with Shibboleth
- Accept and support Unicode
- Have minimal campus branding at the collection and item level

Methodology

The POT1 group decided on an agile approach to the requirements gathering. LT1A was comprised of representatives from five campuses plus the CDL (UCI, UCB, UCM, UCSF, and UCSD). The team created needs assessment questions (Addendum A) and scheduled interviews with all ten campuses. Interview results were compiled, posted to the wiki, and discussed. User stories were created using Google docs for ease of sharing; user stories are simple statements that help distill basic requirements without imposing assumptions, such as “As a contributor I want to be able to upload a single image”. The user stories were further discussed and refined into the critical requirements with priority designation of 1 (highest priority), 2, or 3; as most of the requirements fell under ‘1’, this was removed leaving just 2 or 3 as designations. These requirements were routed to the interview participants for further feedback, and then routed to the POT1 group to verify that the requirements met the expectations of the related Lightning Teams.

The requirements express desired functionality from the campus perspective and are divided into two parts: Content Contribution and Discovery/Display. In some cases, the individual requirements may seem redundant however this is to reflect the nuances of specific functionality. The requirements will be used to help frame the technical modeling and technical requirements and inform iterative development. This document will also help direct policy decisions and should be used as a ‘living’ source document; as the process of building the UC Libraries Digital Collection gets underway, the requirements will continue to evolve and the document updated accordingly.

Interview Participants

UCB: Lynne Grigsby-Stanfill

UCSB: David Seubert

UCLA: Stephen Davison

UCSC: Sue Perry and Robin Chandler

UCSD: Roger Smith and Cristela Garcia Sptiz

UCM: Emily Lin

UCI: Michelle Light and Matthew McKinley

UCSF: Kathleen Cameron

UCR: Eric Milenkiewicz

UCD: Jared Campbell, Daryl Morrison, Sara Gunasekara, Patsy Inouye, Dan Goldstein

Working Definitions¹

Simple Object: a single file or set of related files with no hierarchical relationship between the files; associated with a single descriptive metadata record

Compound or complex object: a set of files with a hierarchical relationship, associated with a single descriptive metadata record. For example, a scrapbook would be a complex object.

Collection: a named grouping of bibliographic items based on some common characteristic, such as provenance or subject. Defined by the contributing campus.

Roles:

- Curator: power user, sets permissions, has contributor access as well
- Contributor: contributes objects and metadata
- Aggregator: campus contributors who have existing digital asset management systems
- Users: those who will be accessing the system through discovery/display only

¹ Source for working definitions: "Features Desired in a Common Digital Library System" created by DISC subgroup for Common Digital Library Platform 6/30/2011.
<http://ufdc.ufl.edu/UF00103112/00002?search=DISC>

Critical Requirements

CONTENT CONTRIBUTION

1. Import/Ingest
 - a. Ingest content files easily via a web interface
 - b. Specify if the content is for:
 - i. preservation and publication
 - ii. dark storage/preservation only
 - c. Persistently manage and store digital objects including, but not limited to, objects with image, audio, video, and text content file formats
 - d. Upload preservation/master copies of content files (e.g., TIFFs for image-based objects) and have derivative/service copies automatically generated when possible (e.g., JPEG or GIF thumbnail images) instead of having to create separate service copies
 - e. Upload individual content files -- or batch upload multiple content files -- with minimal to almost no metadata (e.g., file names only), and subsequently come back and add more robust metadata to the objects

This is to support workflows where scanning or ingesting born digital objects in advance of being able to catalog the resources is necessary
 - f. Metadata embedded in content files to be extracted on import into the system, for ease of working with and adding to the resulting metadata record
 - g. (3) Public contributions will go into a queue so that files may be checked and verified
 - h. (3) Create a virtual space for researchers to upload their own digital objects
2. Content File Management
 - a. Keep local conventions for filenames (for content files), or at least have the system be backward compatible
3. Metadata and Object Modeling
 - a. A web interface for adding/applying metadata to content files, and for editing those metadata records
 - b. The metadata scheme will include, but not be limited to, data elements including: descriptive, technical, and rights information
 - c. The system will support metadata elements that are expressible as simple Dublin Core
 - d. Ability to edit metadata records associated with content files, for individual objects
 - e. Ability to edit metadata records associated with content files, for a batch of digital objects (e.g., search and replace on data in particular metadata elements; replace data across some subset of objects in a given metadata element)
 - f. Upload individual content files -- or batch upload multiple content files -- and simultaneously upload and associate metadata records
 - g. Ability to batch edit groups of digital objects (e.g., search and replace; update data values for a particular metadata element) for quick, global updates to objects (e.g., correct typos, change repository name)
 - h. Integrated use of authoritative vocabulary terms/headings from Library of Congress, Getty, Visual Resource Association, Thesaurus of Geographic Names, or locally created headings within metadata records
 - i. Capacity to describe content in a variety of languages.
 - j. Unicode support for all metadata

- k. Ability to manage access restrictions at the individual digital object level, for example: 1) reading room/department only, 2) library only, 3) specific UC campus only, 4) all UC campus users, 5) specific authorized users only, and 6) public access.
- l. Ability to link campus digital objects back to an external finding aid or collection description, through the metadata
- m. Ability to associate one or more digital objects with a collection (defined by provenance, topic, etc.), through the metadata
- n. (2) Structure metadata records-to-content file relationships as simple digital objects, such as photographs
- o. (2) Structure metadata records-to-content file relationships as complex digital objects, such as scrapbooks and multi-page items, files/aggregations of materials, oral histories with texts and audio, etc.
- p. (2) Ability to easily generate metadata exports in other standardized XML formats
- q. (2) Easily generate a metadata export in a simple format, such as CSV or tab-delimited files, for local re-use purposes
- r. (2) Metadata and digital objects exposed as an OAI service; metadata fields TBD.
- s. (2) Ability to select and describe content from across multiple collections, creating artificial, topical collections within the local institution and across the UC system.

4. Publication and Access Control

- a. Allow for publish and unpublish of a digital object; unpublished digital objects would only be available for viewing to curators and staff (catalogers, etc.) using the system to create and manage objects before making them publicly available.
- b. Publishing of digital objects with a quick turnaround time for availability to end users (e.g., semi-immediate or <24 hours), to be responsive for faculty requests for objects, users with scan-on-demand requests, etc.
- c. Control the degree of user access to a published digital object, based on the access level designated in the metadata record for the object.
- d. Allow for change to access controls (and resulting publication status) for digital objects in real time, in response to a copyright complaint or when an embargo ends.
- e. Limit what is shown to end users at different access levels; for example, show only the metadata record for a given digital object -- but not the content file(s).
- f. Ability for end users to discover a specified repository's digital objects -- e.g., a browse or limit by (UCB, UCSF, UCLA, etc.) repository -- for ease of pointing users to local items.
- g. Published digital objects to display full metadata records, so users have complete descriptive, rights, etc. information for an informed use of the resource.
- h. Published digital objects to indicate what collection(s) they are associated with.
- i. Published digital objects clearly branded with the institution name, so it's clear to end users who/where to contact for more information, which institution manages the content, etc.; URL branding is less important.
- j. (2) Ability to associate campus digital objects with thematic or topically-clustered curated collections in the UC Libraries Digital Collection access layer.
- k. (2) A mechanism that monitors embargoed digital objects; when embargo ends, the curator is notified for further action (e.g. publish, extend embargo)
- l. (3) Access control on an item for a short time for things such as ILL.

5. System Administration and Management
 - a. A centrally hosted solution
 - b. Access to viewing and browsing digital objects at the campus level (e.g., see thumbnails or service copies) to quickly identify objects that a campus is working with.
 - c. Define user accounts with different permission levels (e.g., full admin privileges, metadata editing only, etc.), for different staff that will be creating digital objects (e.g., student assistants, etc.)
 - d. Ample digital asset storage space for each campus (allocation to be determined)
 - e. Access to user statistics (e.g., number of hits, terms queried, location)
 - f. Access to campus repository statistics (e.g. number of items in a collection, number of collections)
 - g. (2) Ability to assign different parts of the workflow to internal experts (e.g., metadata to a metadata librarian)
 - h. (3) Ability to track hours spent on metadata creation, preservation, and other workflows.

AGGREGATION

Assets harvested from existing campus-based DAMS into the UC Libraries Digital Collection

1. Easily access metadata for harvesting.
2. Identify which records from a campus were added during last harvest.
3. A set of API's to interact with the repository/repositories where the digital objects are being managed.
4. A persistent reference for each UC campus.
5. A persistent identifier assigned to each object at ingest, for end user access.
6. The repository should appear in digital object metadata records (e.g., URI or identifier) so campuses can consistently manage and track information for the repository for the purpose of generating reports, statistics, etc.
7. Harvest digital objects using an OAI interface in order to gather all of the metadata.
8. Clearly identify that content held in federated repositories is from the contributing institution.
9. Relate objects to non-provenance based collections.

DISCOVERY AND DISPLAY

The requirements below reflect information gathered during the interview process and from a range of existing requirements documents authored by UC campus library staff.

Search

1. Basic search: Every page should include a single text box for simple keyword searches that may include single or multiple search terms. When a keyword search is submitted, the following fields will be searched: title, subject, description, contributor, date, format, rights.
2. Advanced search: All metadata fields exposed in the search results display should be available to be searched independently or in combination from an advanced search page. The exposed metadata fields will include all available fields in a given metadata schema.
3. Scope: By default searches should be conducted across all collections with the option of limiting to a specific collection.
4. Spelling correction: Search term spelling correction should be provided.
5. RSS: Users should have ability to subscribe to RSS feeds in lieu of stored queries.
6. Multilingual search: Search should accommodate multiple languages. Unicode support.
7. Search engines: Content should be optimized for and discoverable via search engines.
8. Mobile devices: Content should be discoverable and displayable via mobile devices.

Search Results

9. Item level information: Each item in a result set should be accompanied by the following primary metadata: title, subject, description, contributor, date, format, rights.
10. Facets: Facets should serve to refine or expand search results and should be made available for the following primary metadata: title, subject, description, contributor, date, format, rights .
11. Sorting: Default sorting of search results should be by relevance; users should have option to sort by additional sorting criteria: collection, author, title, date.
12. Items per page: Users should be provided option to display pre-set items per page (e.g., 10, 15, 20).
13. Pagination: Result sets should be paginated with users able to navigate back / forth through pages of results.

Object View

14. Context: Objects should be displayed in a view that provides UC Libraries Digital Collection, UC campus, and potentially collection-branding.
15. PDF display: PDFs should be displayed within the branded area and not in a separate Adobe Acrobat Reader window.
16. Thumbnails: Images should be represented by thumbnails that when clicked open to a full view of the image within an image viewer.
17. Image viewer: Images should be easily optimized for viewing, including zoom in/out, rotate, mirror/flip, fit image, and full size.
18. Search terms: Search terms should be highlighted in the object view, regardless of format.
19. "More like this": Items similar or related to the displayed object should be linked to from the object view page allowed users to view "more like this".

20. Object level citation: All objects should have an object-level citation. A “Citation” link or icon should be available that when clicked will display citation information.
21. Download: A link or icon should be available on all object views that when clicked will allow the user to save the selected content.
22. Print: A link or icon should be available on all object views that when clicked will allow the user to print the selected content.
23. Purchase: A link or icon should be available on all object views that when clicked will provide users with the contributing institution’s contact information.
24. Item / book bag: Users should be able to click a link associated with each object to add a citation and actionable URL to a session-based item / book bag page.
25. Email item / book bag: Users should be able to email to themselves or others the objects saved to a session-based item / book bag page.
26. Social media: A link or icon should be available that when clicked will allow the user to send objects to social media targets (e.g., Facebook, Delicious, Pinterest).

Attribution

27. UC Libraries: The UC Libraries attribution/brand should always be present; all pages should have a branding area at the top that will include at minimum the UC Libraries brand.
28. UC campus: UC campus attribution/brand should be present on all pages associated with the contributing campus.
29. Contributing institution: Objects contributed by or associated with a given entity will be identified on the object level page in the area containing associated primary metadata.

Feedback / Communication / Inquiries

30. Help / feedback: A link or icon should be available from all pages that when clicked provides a feedback form for submitting comments and questions to the UC Libraries Digital Collection staff.

Contributor / Collection Information

31. Contributing institutions: A full alphabetical list of contributing institutions should be made available, with each entry linked to a customized landing page including full contact information. The right to perform administrative activities relative to the landing page (e.g., institution contact information) should be granted to the contributing institution.
32. Collection description: A document describing the collections included in the UCL Digital Collection should be available on the site.
33. User guides: A document describing how to use the features the UCL Digital Collection should be available on the site.
34. Contributor guide: A document providing guidance for how to contribute to the UCL Digital Collection should be available on the site.
35. Technical documentation: A high level description of the components driving the UCL Digital Collection should be available on the site.

Addendum A: Interview Questions

Context	Questions
Understand current work and capacity	Who at your campus is actively participating in digital content/collection creation? ex. data curators, digital librarians, tech services, metadata librarians, special collections, IT staff.
Workflow	Describe your current workflow. (selection, ingest, metadata, access, preservation, etc.)
	What kind of content do you primarily work with?
	What is the most complex digital asset type you work with?
	Describe your most complex collection
	Do assets always belong to collections, or are there one-off assets? Describe the type of hierarchy used to organize digital assets.
	What's your workflow for creating metadata, and loading and staging files associated with the metadata?
Management	What kind of system are you using to create and manage your digital assets, and what features are you using?"
	What are your file naming conventions, if any?
	What are your language requirements for creating records?
	How are you utilizing and managing authority records and controlled vocabularies, within the context of your system?
	In addition to managing digital instances of resources, do you also need to manage information about physical/analog assets (in cases where they've been digitized) such as location information, loans, etc.?
	Are you managing end-user contributed content in your system?
	What kinds of export formats can you generate (or need to) for your collection?
	How are you preserving your digital content?
	Do you have a local means of providing access to your digital collections, and is it satisfactory?
End user needs	Who are your key users?
	How and where do they engage with digital assets? What features are important to them?
	Do your users have any specific language requirements?
	Do end users need to maintain "personal collections"? such as bookbag or a shopping cart?
	Are mobile users part of your desired user base?
	What kind of search functionality do your users need in an interface? What do you need?
Shared system	How would you prefer to submit digital content and metadata to a shared service?
	Are you able to provide metadata, and, if so, what kind?
	What kind of single item asset editing features do you require? Bulk editing?
	Will assets need to be shared between collections? Please describe how you would utilize this.
	What access control mechanisms are needed?

	What would make it easy to contribute to a shared collection? What would make it most difficult?
	Is it important for you to include metadata records and/or links to your digital objects in Melvyl?
	Where are you currently sharing your digital objects? OAC, HathiTrust, Internet Archive, Web Archiving Service, etc
	Are there other programs/program interaction that you would like?
	Are you concerned about rights management? Would a copyright guide or "calculator" be useful?
	What kind of reporting features/functionality do you require?
	Assuming you want statistics, is Google Analytics acceptable to use? Are there other reports needed?
	What is your total digital asset storage requirement now and in 5 years?
	Are there any plans for the future for your digital library resources that we should know about?