Functional Requirements for a DAM: The William Andrews Clark Memorial Library

Description of Site

The William Andrews Clark Memorial Library is a UCLA off-campus, non-circulating library located on a historic five-acre property in West Adams that had been bequeathed to the university by its former owner William Andrews Clark in 1934. The library specializes in English rare books and manuscripts dating from the Tudor period to the long eighteenth century, as well as materials from the rest of early modern continental Europe. Though originally part of the UCLA library system, the Clark is now administered by the Center for 17th & 18th Century Studies. From a technological standpoint, the library is distinctive in its continued use of the card catalog (as a supplement to the regular UCLA-wide online library catalog) and its lack of a workflow management system for requesting materials. As a Los Angeles Historic-Cultural Monument, the ground's features (furniture, facilities, and artwork) are not to be altered, and must be preserved accordingly to the appropriate HPOZ (Historic Preservation Overlay Zone) guidelines. The Library currently receives grants from institutions such as the NEH and CLIR to perform large-scale digitization projects of their early modern manuscripts and annotated books, many of which are currently hosted on Calisphere through a partnership with the UCLA Digital Library.

<u>Staff Members Interviewed:</u>

Anna Chen	Philip S. Palmer	Scott Jacobs
Head Librarian	Research Services	Reading Room Supervisor
achen@humnet.ucla.edu	crobinson@humnet.ucla.edu	sjacobs@humnet.ucla.edu

Alejandro Sanchez Nuñez	Carole Robinson
Cataloging Assistant	Site Manager
sanchez@humnet.ucla.edu	crobinson@humnet.ucla.edu

1. USERS

The hypothetical DAM system will primarily be utilized by the Research Services Department of the library (composed of 1-2 people), which will be creating most of the digital assets to be stored in the system. The digital assets emerging from this department, generally high quality scans of entire rare books and manuscripts, are generally amongst the more valuable digital assets of the library, resulting from projects funded by institutions such as the NEH and CLIR (Council on Library and Information Resources). Currently, the scanning is performed by the third-party service Luna Imaging, with the high-resolution scans of the documents being sent back to the Clark Library on a single hard disk drive (no redundancies, though the images are backed up on tape at the imaging center) that is then taken over to the Digital Library, where the

files are ingested into Nuxeo and uploaded onto Calisphere. The Head of Research Services has expressed that the Library's use of Calisphere was a decision born out of convenience rather than it being an ideal fit, and that they would be open to an alternative system that would grant them more flexibility with what could be done with the assets. Research Services also handles quality control of the assets while the items are being scanned and after they have ingested into the system. checking to see if the metadata matches up with the physical items. Some metadata management may also be performed by those in the Cataloging department.

The second group of internal users will be those in Reader Services (an operation run predominantly by one person), who seek a DAM system to manage reprographic requests from a variety of sources, ranging from museum institutions to individual researchers. Reprographic requests are fulfilled using simpler means – materials are usually photographed using a camera, rather than professionally scanned. Currently, all the assets for this department are stored on the computer of a single staff member (who serves as the asset's main point of access). The DAMS system would allow this individual to better keep track of which assets have already been created, as well as allow them to retrieve requests more quickly. With the acquisition of a DAMS, Reader Services has posited the possibility of making a greater variety of assets (i.e. video and audio content).

The third internal user group is the Site Manager, who believes it would be beneficial for the library to have a system that could be used to track the conditions of the library's furniture and other decorative features (i.e. artwork). The DAM system would help the library keep track of such assets according to the Clark's needs as a historical preservation site. For example, images could be taken of the furniture, placed into the DAMS, and be used to track upholstery changes overtime.

External users include the staff at the Center for 17th & 18th Century Studies, who will need to find the necessary materials for predominantly promotional needs, but they will primarily by interacting with the DAM through the content managements system that it is integrated with, although. Public users – the library patrons and other scholars – will likely interact with the DAM in a similar fashion.

The putative DAM system will have to support a minimum of 3-4 internal users (with differential levels of access) and be able to integrate with a CMS for public access.

2. FORMATS

The Clark's assets are predominantly photographic, although, as mentioned there are plans to include a greater variety of assets should there be the existence of a system to manage

them all. Currently, the files for the digital projects are distributed across several hard disk drives, with copies existing on the servers for the Digital Library, so size estimates are hard to make. Based on rough estimates, as well as a survey of the library's shared drives:

Current file formats	File storage (in GB)	Estimate 5-year growth
TIFF	3500	variable, depending on project
JPEG2000	500	100
PNG	400	80
CR2	< 1	< 1
PDF	300	50
.psd (Adobe Photoshop Image)	10	< 1
.doc	< 1	< 1
.xlsx	< 1	< 1
MP4 (H.264 codec)	< 1	50

Total storage size: ~4500 GB, ~4.5 TB
Estimated growth in 5 years: 0.5-1 TB

3. CONTENT

The DAM system will primarily be used to organize the following digital asset contents:
- Professional grade high-resolution pages of early modern books and manuscripts
    o (~85%)
- Digitized materials (manuscripts, parts of books, illustrations, and images from the Clark Library collections aggregated from a history of reprographic requests)
    o (~10%)
- Facilities management (photograph of the grounds, repairs, furniture, artwork)
    o (~1%)

Potential secondary assets for inclusion in the DAM:
- Programming (i.e. invitations for speakers, conference materials, itineraries, hotel checklists, anniversary events)
    o (< 1%)
- Library workflow documents (i.e. acquisition information and collection files, purchase receipts, correspondence, deed of gifts, loans)
    o (< 1%)
- Photographs (i.e. from exhibitions, events)
    o (2%)

- Promotional materials (i.e. event flyers, newsletters)
    - (< 1%)

Looking forward, the library is considering the creation of short documentary-type videos regarding books and authors that are prominently featured in their collections.

4. CONVENTIONS

Currently, file-naming conventions are employed inconsistently across the different asset types and are mostly exclusive to the creator of the file. Depending on the asset, there is usually some indicator for the date of the content matter (i.e. the year a copper-plate engraving was made) in the asset as well as a descriptive statement.

- Sample naming conventions:
    library_gatehouse_history_1987
    MS_2015_014_bookseller_description_item39
    Bruman_Payment_iPalpiti Soloists_2 Performance Payments
    UCLA_Clark_Library_10.24.13
    Collection#_DescriptionOfItem
    uclaclark_call#_page#

Assets for Digitization Project

The file naming conventions for assets produced as a result of the digital projects have a decent file naming convention: uclaclark_call#_page#. One element to add may be the project ID under which the asset was made, if this information is not already made available in the metadata.
    Example: uclaclark_projectID_call#_page#

Since the call number that comprises the bulk of these filenames is oftentimes long and convoluted (for example "QL673.A31779*"), it might be beneficial if the DAM system could have some system in place to check that the correct call # was entered in its entirety into the filename. Within the catalog, the call number of a lot of the rare books have been appended with an asterisk to indicate that it is a rare book. In such cases, the asterisk should be removed since asterisks have their own meaning as command functions in many DAM systems, and be replaced with a different code like "RB".

Assets for Reprographic Services:

Currently, the file names follow the schema "Collection#_DescriptionOfItem". It is recommended that some extra information be added in regards to its physical location (i.e. box and folder number), and that the description of the item be kept in a separate spreadsheet. For reprographics of archival materials, it might be beneficial to include information such as the box and folder numbers. For books, the call number and a page number should be included. If edits have been made to a particular file, they should be indicated with information regarding the version number; the master copy should be clearly labeled and kept as well. Same as with the files for the digitization project, a way to easily cross-reference the materials in the DAM to the library's existing catalog records would be helpful as well.

Assets for Furniture/Artwork Management

Furniture inventorying currently utilizes unique IDs (termed property IDs), which are physically noted on the furniture. Photographs taken of the furniture have their own separate numbering system: a simple "001, 002, 003…" numbering scheme that seems to reflect the order in which the photo was taken. Since a lot of the furniture has yet to be systematically catalogued/identified, the use of these two IDs (property IDs and photograph IDs) persists, and the two different numbering systems are corresponded with the item description using an Excel spreadsheet. Upholstery changes are also not tracked in an organized manner. The artwork at the Clark is currently being catalogued and will be assigned unique bibliographic numbers and call numbers based on their location. Within the DAMS, the file names for the photographed/scanned artworks should include information on who created the artwork, its title (if available), and information regarding the digital version made of the work (publication-grade, low resolution)

The DAMs system could help to streamline this task, keeping track of multiple numbering schemas as well as manage "version control" resulting from different manifestations of a piece of furniture throughout the course of upholstery changes. It is advised that the photographic files of the furniture – those with the numbering system "001, 002, 003" – be changed to a more descriptive scheme that includes information such as the type of upholstery work that has been done, a date of when the photograph was taken, and a brief description of its location. A controlled vocabulary should be created for the different types of upholstery work.

Example: photograph#_upholsteryWork_mmyyyy_location

If the item has been properly identified and assigned a property ID, the photographic manifestation of the piece of furniture should be given a filename that reflects this, instead of the "001, 002, 003..," numbering system.

Example: propertyID_upholsteryWork_mmyyyy_location

General Suggestions:
- Regardless of the format/content of file, avoid the use of special characters in file names, in particular: " / : * ? ". Use underscores instead of spaces to separate elements when naming files

- Indicate version types in file names and keep original master copy (should be the one with the best quality)
- For any video related assets that emerge in the future, here are some potential file naming conventions:
- If the video is regarding a specific book/manuscript in the Clark collection
Example: VID_DescriptionOfVideo_Call#OfBook
- If the video is regarding an individual, use their standard names as listed by ULAN (Union List of Artists Names)
Example: VID_DescriptionOfVideo_ULANentry

5. METADATA

The Library's digital assets are currently described at varying levels depending on the group of assets that they belong to – digital projects, reprographics, or site management. The reprographic materials and photos of site facilities/furniture do not need to be described to the extent as the materials produced for formal digital projects, although the artwork is formally cataloged. Since there will be different types and extents of metadata, the DAM must support customizable metadata fields.

Assets for Digitization Projects

Currently, the metadata for the rare books and manuscripts can be understood as falling under "item level" and "page level" metadata. Page level metadata is currently being entered into an Excel spreadsheet, and quality control is performed to make sure that the data in the spreadsheets correspond with what is uploaded into Nuxeo at the Digital Library. Item level metadata (referring to records for the entire book or manuscript) can be found in the Voyager cataloging system and is publicly accessible through the UCLA Library catalog. One important feature for the Clark's own DAM would be to have a straightforward way for item-level metadata in the cataloging record to be transferred over. The MARC records would have to exported in MODS/XML before they can be ingested into the DAMs. The metadata for the materials can sometimes be in languages that utilize diacritics, and there are oftentimes special symbols (superscripts, subscripts, double daggers ‡, sections §, etc.) that are used to describe certain aspects of the book such as their signatures and pagination. As such, the DAM system should also support Unicode. The DAM should also ensure that metadata for the assets in this category can be mapped onto DublinCore, with such information made exportable at the front end (CMS).

Assets for Reprographic Services:

The individual managing reprographics utilizes a local metadata schema consisting of a relatively limited number of fields:

- collection number
- item's year of publication
- page number
- document type
- a brief title describing the content (i.e. "Oscar Wilde American Tour")

While these assets do not warrant extensive technical metadata extraction, it would be helpful to include information about the file resolution, as well as some metadata regarding the process of asset creation (i.e. who scanned the photo, who made edits to it). Since many of the assets are requested for publication elsewhere, some administrative metadata (addressing issues of copyright and licensing) would also be necessary. Having a formal system through which public users can make requests for items could also help facilitate the process of monetizing reprographic services.

For a time being, there was an attempt to use Adobe Bridge to attach metadata onto an asset, but the entire process turned out to be rather cumbersome as there were some difficulties performing bulk ingest. Reinstating the use of Bridge to manage new assets created could serve as a short-term solution and could probably ensure that the assets are better prepared for transfer into a formal DAM system.

Assets for Furniture/Artwork Management

The furniture is described based on a local metadata schema including information regarding its:

- location
- dimensions
- creator
- monetary value
- physical description
- provenance

There is no need for technical metadata extraction with these assets, as they will only be for internal use and to track upholstery changes and locations of the furniture within the library.

Regarding the artwork, the DAM system should support existing thesauri such as the AAT (Arts and Architecture Thesaurus) as well as a local thesaurus creating tool. Similar to the digitized rare books/manuscript assets, the artwork will have a cataloging record. Information in the catalog record should also be present in the DAMs, with the addition of fields for describing specifics for the digital manifestation of the art:

- Resolution of file
- Date when photographic image was taken
- Conservation status and preservations changes (when relevant)
- Usage in exhibitions

Overall, given that many of the digital assets will relate in some way to a singular physical item (a book or a collection), the ideal DAM system should support parent/child

relationships where all the different manifestations (i.e. different scanned versions of a photograph or the pages of a book) can be linked back to a parent work (the book or archival collection). Most of the metadata for the assets is encoded in MARC, so the putative DAM system should be one that can take in XML-rendered information from the original MARC records. Since much of the metadata is also currently being entered into an Excel file, the Clark would need a system that can take in CSV files.

6. SEARCH

The putative DAM system will need to support faceted and keyword searching. Oftentimes, patrons make very specific image requests and faceted searching based on a specific metadata field would make it easier to find items. Other times, external/public users may request images based on subject tagging (i.e. "alchemical imagery"), so general keyword searching is also desirable. A sizeable portion of the assets, particularly those resulting from early 2000s reprographic requests, has little to no metadata attached, so the DAMS should support some manner of visual recognition, OCR (optical character recognition), or HTR (handwritten text recognition) to make them findable if retroactively attaching metadata is no longer feasible.

7. DISPLAY AND ACCESS

Currently a chunk of the library's digital assets is all on a shared drive that is close to capacity, making the folders and files within the drive hard to open. The DAMS should support the ability to preview assets without having to open the entire file, as many of the staff members have remarked that opening the files takes up a sizable portion of the time spent on searching. All assets should be downloadable (and with future inclusion of video assets, streamable). With regards to the digitized book and manuscripts assets, materials should be downloadable at both the parent (entire book) and child level (individual pages).

Public access to the Library's digital assets, asides for the digital projects hosted on Calisphere, is all currently mediated through email or phone requests with one library staff member. Through the CMS, public users should be able to make requests for certain assets, which can be approved by the appropriate staff member.

8. WORKFLOW

Assets for Digitization Projects

     After being scanned by a third-party imaging vendor, the files will be bulk-ingested into the DAM system, where they can be identified as being in a "pending" status. The system will also take in the appropriate metadata from exports of library MARC records, either from CSV or XML source formats, and allow for bulk-metadata editing. After the appropriate quality control measures are taken (comparison with the physical item), the status of the assets can be changed to "approved". Digitization project assets are intended to be free for public use, and such users should be able to download entire copies of books/manuscripts, as well as metadata information by requesting from the CMS.

Assets for Reprographic Services

     After the images are photographed on-site, they will be ingested into the DAM system. The DAM will facilitate the bulk conversion of the assets into 3 different image file types – TIFF, PDF, and JPEG2000. When assets are requested, the user should be able to easily search for relevant items based on a specific metadata field or general keyword search. Since the items would be in a database accessible to a number of users, the task of answering reprographic requests could be distributed. Through the DAMs, the staff member should be able to issue the client a link to the file (through which they will be able to download the asset). The DAM system should cut down on the creation of redundant files and make it easier to manage copyright and intellectual property issues if relevant.

     There is currently a sizeable backlog of digitized collection materials from staff members who have long since left the organization. Since many of these assets come from the same collection or deal with similar themes, a way to bulk ingest files and to perform bulk editing would significantly expedite the processing of the backlog. As mentioned previously, visual recognition, OCR (optical character recognition), or HTR (handwritten text recognition) could perform auto-tagging of such assets and make them searchable even with limited metadata.

Assets for Furniture/Artwork Management

     Using the DAM system, the site manager will be able to create assets that help track preservation changes to furniture/artwork and link photographs to the physical object and their locations. The system could also help to manage the status of certain art pieces, which may sometimes get loaned to other institutions.

9. PUBLISHING

The Library is interested in a content management system that would help them push their assets onto a public platform – for example, a CMS such as Omeka. With future video assets, it might be necessary to publish them onto a platform such as YouTube or Vimeo.

10. DIGITAL PRESERVATION

The Clark's digital assets are not currently preserved in a systematic way, although there could be a policy in place for the materials that were sent to the Digital Library. Particularly regarding reprographic requests, a lot of the files made preceding the mid-2000s have become largely obsolete or have been corrupted with time.

The organization will need to have a DAMS that automatically generates MD5 checksums, although such checksums do not seem to have immediate use to the organization's day-by-day operations (at least, not given in current procedures).

Redundant copies of the hard drives containing files for the digitization project should also be made, particularly if the Clark is considering moving those assets off of the Digital Library's servers in the future. Since the Clark's current servers are shared with the Center, it may be beneficial to have s SaaS system that handles storage through cloud-based servers and conducts preservation actions.

11. USER RIGHTS AND SECURITY

All files are currently on a shared drive between the Center for 17th & 18th Century Studies and the Clark, with everyone having basically the same access and editing rights. Historically, this has lead to files getting improperly edited or deleted, as well as numerous cases of creating redundant copies of an asset. The DAM will need to support different levels of access, with some people having the rights to create, edit and view the files, and others to only view them.

12. INTEROPERABILTIY

The proposed DAM will need to support International Image Interoperability Framework (IIIF) functionality, particularly the Image API and Presentation API. The DAM system should also interoperate with a CMS.

13. TECHNICAL

The Clark library utilizes both the Windows and Apple platform. Since the Library is unlikely to hire a programmer to create a tailored instance of an open-source DAMS, it may be necessary to seek out a proprietary SaaS DAM services. This would also help with storage concerns.