Initial Report of the Digital Asset Management Assessment Task Force

Task Force Members

University Libraries
Chair: Sarah Potvin (Office of Scholarly Communication)
Robin Sewell (Medical Sciences Library)
Julie Mosbo (Preservation)
Pauline Melgoza (User Services)
Jeremy Huff (Digital Initiatives)
Lisa Furubotten (Cataloging)
Anton duPlessis (Cushing)
James Creel (Digital Initiatives)
Sean Buckner (Preservation)
John Bondurant (Cushing)
Greg Bailey (Cushing)

College of Architecture
Ian Muise (Technical Reference Center)

Marketing & Communication
Brian Wright

Sponsor
Paula Sullenger (Information Resources)

April 21, 2016

Overview

Charge

In fall 2014, a DAMS assessment task force (hereafter "TF" or "we") was established with the objective of investigating and making recommendations for a solution or solutions that would enable the Texas A&M University Libraries to store, display, and preserve new forms of university information and research. In the spring of 2015, the charge expanded to include attention to broader campus needs.

Process

After defining an assessment process and expanding our scope to include campus, the TF first worked to conduct a campus needs assessment, to identify and develop use cases, and to distill core requirements. This became the basis of our testing rubrics. We ran multiple stages of assessment to identify and test systems, as well as to analyze the results of those tests. A recommendation was reached on the basis of this analysis, and further inquiries. For a full list of activities covered under this process, see "Process" section.

Executive summary:

Our analysis of twenty-six systems allows us to confidently assert that no one digital asset management product will meet library and campus needs. Broadly, "digital asset management consists of management tasks and decisions surrounding the ingestion, annotation, cataloguing, storage, retrieval, and distribution" of image, multimedia, and text files.[1] These tasks are performed by systems (DAMS) that differ in their approach to functions and range of associated capabilities. Given campus needs, and our experience with the DSpace DAMS, the task force was attuned to the particular importance of the data models embedded in these systems, which guide and constrain other functionality.

We are convinced that modular solutions to discrete needs for storing, displaying, and preserving digital assets are warranted, and that these solutions are likely to require customization. We recommend building a digital asset management ecosystem (DAME) rather than attempting to meet all needs with a single DAMS. In addition to more granular tradeoffs among the DAMS systems that we consider, there are **two main points of decision** for a DAMS component that we put forth:

- Our reliance on local development versus vendor development/deployment;
- The deployment of relatively full-featured but potentially inflexible systems versus incremental/iterative deployment of more modular, flexible systems.

[1] https://en.wikipedia.org/wiki/Digital_asset_management

Dependencies & Unknowns

At this stage, we assume the continued deployment of DSpace and Archivematica and the forthcoming deployment of VIVO and the Texas Digital Library's Dataverse as complementary open-source systems, and our analysis considers the need to provide a solution that will bridge the gaps. DSpace provides a fully-featured solution for institutional repository needs, including format-agnostic file storage and delivery, extensible flat metadata, integration with Archivematica, and robust permissions. Dataverse is expected to provide targeted support for data, with functionality for description, permissions, the generation of data citations, and publishing workflows. Both DSpace and Dataverse boast APIs and UIs. Archivematica, our digital preservation system, currently supports limited integration with DSpace. Its developers continue working to make the program as interoperable as possible with other content systems, with future plans to integrate with Fedora, Islandora, Hydra, and Dataverse, among others.

Governance for a campus solution has not yet been determined: it was understood that this process would be further defined once the Libraries made a determination.

Also outside of the scope of the TF was recommending specific exhibition systems or workflows and policies. A subsequent task force, with different composition, is expected to recommend policies. An exhibition sub-group is currently looking at those systems.

The systems that the TF considered are constantly evolving; the functionalities outlined below may have changed or be in the process of changing. At the time of this report, Hydra-in-a-Box, an IMLS-funded effort directed by the Digital Public Library of America, Stanford University, and DuraSpace, was still in project design phase and had not yet been released. Hydra-in-a-Box has been advertised as an out-of-the-box next-generation digital repository, "useful to any institution managing collections of digital assets," it is expected to build on the Sufia 7 Hydra head, which sits atop Fedora 4 (note, below, that the TF tested Sufia 6).

Our Recommendation

The TF recommends the deployment of modular digital asset management components to meet the complex needs of the Texas A&M University Libraries and campus. These include:

- The deployment of a system to manage and store digital assets and metadata. Our recommended open-source system is Fedora 4, to be coupled with Blacklight and Solr for search and retrieval. Solr indexes content managed by the repository, and Blacklight enables search and retrieval across the indexed content.
- Nuxeo may serve as a vendor-supported alternative to the Fedora-Solr-Blacklight stack. This fully-featured enterprise DAMS is more fully described in the next section.
- The development of custom user interfaces as appropriate (likely, public user interface and administrative interfaces).

- The deployment of a triple store to enable linked data, along with Apache Camel and Fuseki as the basis of connecting Fedora to the triple store and to Solr indexing software.
- An exhibits system, to be determined by the Exhibits Sub-Group.

If Fedora 4 is selected, the TF also anticipates the need to develop or deploy metadata creation and editing tools.

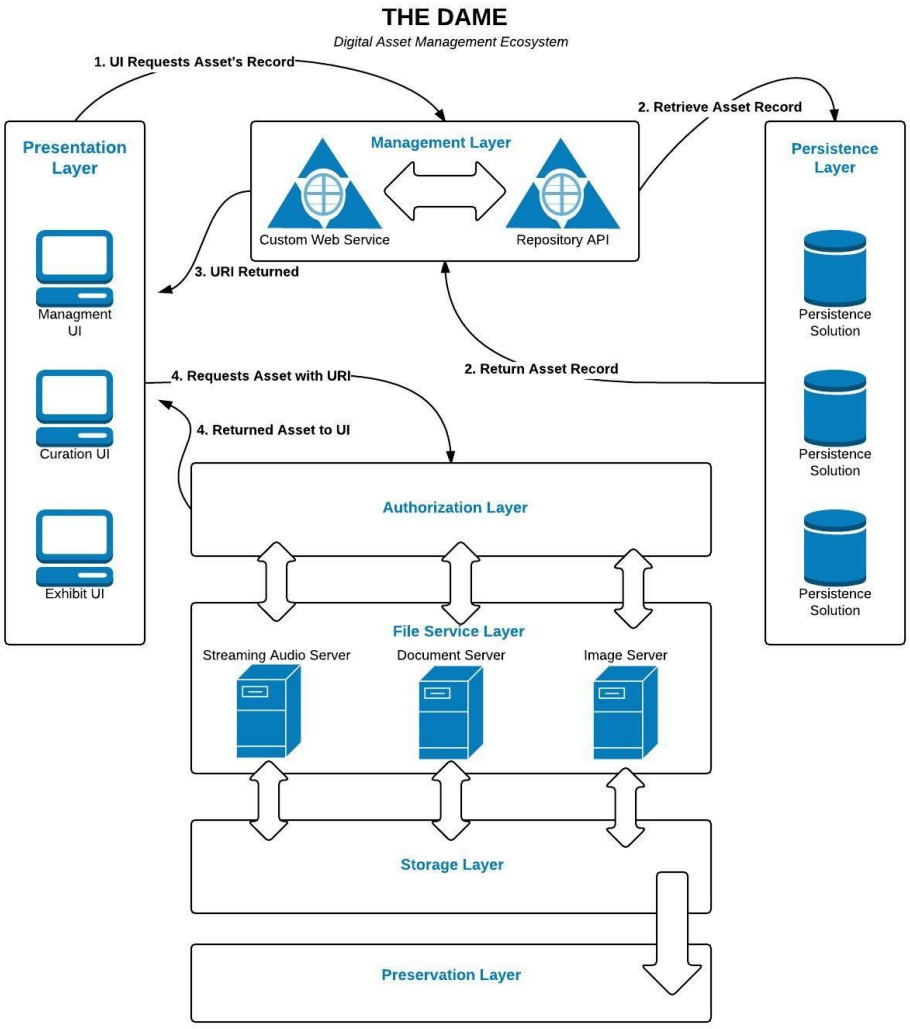
Because Fedora functions as a component rather than a standalone DAMS, it was not separately tested. Fedora is the repository underlying both Islandora (currently Fedora 3) and Hydra/Sufia (Fedora 4), so it was tested both directly and indirectly in the course of testing those systems. As a repository, Fedora does not suggest immediate solutions for user interface, workflow, or data structure out of the box; institutions must decide solutions to these areas locally. The fact that Fedora does not address areas like UI is not necessarily a shortcoming: by not imposing a locked-in solution to these problems, Fedora 4's modular approach allows us to be flexible.

A Note on Terminology:

Digital asset management is, by its nature, technical. We have attempted to use approachable language where possible, and to define terms and acronyms where they first appear; for further information, consult the glossary that precedes Appendix 1.

Our recommendation is to emphasize building a digital asset management ecosystem (DAME) over selecting a DAMS. What is the difference between these two approaches?

In a pure sense, a DAMS is the management layer that links item/metadata storage to user interfaces (UIs) and administrative tools. In our DAME schematic, the DAMS functions to communicate and direct actions between these areas:



The DAME is a Digital Asset Management Ecosystem. The choice of the word ecosystem, as opposed to “system” (as with a DAMS) is explained by the DAME’s emphasis on a distributed service architecture. This is an architecture in which the discrete roles of a DAMS are handled not by one application, but instead by a collection of applications, each one suited for the role it plays. The DAME’s structure will certainly vary from institution to institution, and in fact this flexibility is perhaps the DAME’s strongest quality. In general, a DAME’s ecosystem will be divided into the following layers:

- Management
- Persistence
- Presentation
- Authorization
- File service
- Storage
- Preservation

Each of these will need to be addressed individually.

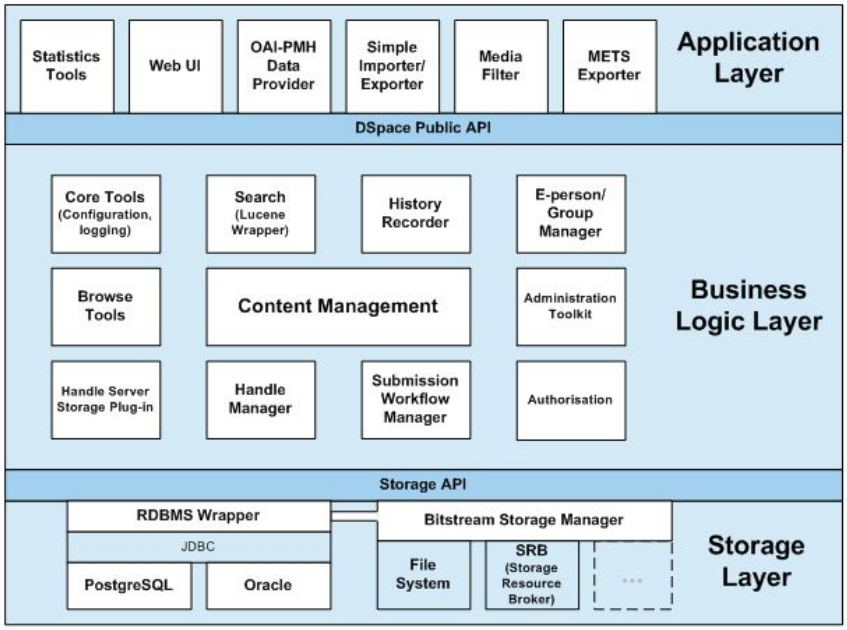
In the DAME, the management layer is conceived of as a collection of web services that handle record creation, curation, and discovery. It does not, itself, handle the actual assets, but instead records the assets' location and metadata, and allows for the management and retrieval of this information. The management layer should be comprised of at least two elements, the first being a custom web service and the second **a repository with a fully featured application profile interface (API). The repository application can be one of the many popular DAMS solutions that are currently in use, the only requirement being that it exposes all desired functionality through an API.**

It may seem that a repository with a fully featured API would be sufficient to satisfy the needs of a management layer, but there are several good reasons for including a custom web service in this layer. The first reason is that this webservice will act as an interface for all communication with the management layer, and by so doing, the DAME is repository agnostic. All other applications in the ecosystem will be programmed against the consistent API of the custom service, and the job of interfacing with the repository's API is left solely to the custom web service. If the decision is made to switch repositories, the only thing that needs to be updated in the DAME will be the custom web service, and the rest of the ecosystem will not realize the change took place. The second reason for this separation is its allowance to employ multiple repository solutions side-by-side, with the web service aggregating their responses. Finally, in record retrieval, the authorization and authentication of the user can be handled by the custom web service, relieving the repository of any need to be compatible with the institution's authentication and authorization strategy.

This management layer thus communicates with the persistence layer, which is not, by necessity, one of the more complicated portions of the DAME's architecture. It is simply the data source, or collection of data sources, needed to support the repository. Most repositories that would work well in the DAME are likely to have varied options when it comes to persistence, making the persistence layer one of the more flexible aspects of the DAME. In general this layer will store the assets' URI, metadata, and possibly even application-specific information needed by the presentation layer.

It is perhaps worth emphasizing that the DAMS described in this report typically consist of many tools and systems that interact with one another as layers, containing several of the layers envisioned and described in the DAME. For example, DSpace is typically described as a three-layer system[2]:

[2] See "DSpace 5.x Documentation: Architecture," https://wiki.duraspace.org/display/DSDOC5x/Architecture



Every system that we tested consisted of different tools and components, bundled together as a single system. Part of the argument for a DAME over a DAMS is the ability to determine the components in these bundles locally, and to swap them out to meet evolving needs.

Opportunities with Digital Asset Management:

Technological advances continue to improve the way digital repositories handle, manipulate, and display collections beyond presentation of simple files and streaming media. Book readers that allow patrons to page through manuscripts are becoming more common, GIS data and maps are being used to access information in other formats, and there is increased interest in production of 3D exhibits that allow virtual patrons to handle and examine three dimensional artifacts. The semantic web and linked data have the potential to integrate information from multiple open repositories to create a unique and personalized learning experience. These methods of content manipulation and delivery require a robust user interface as well as access to complex and multilingual metadata schema, indexing, and linked data. A DAME architecture would provide the necessary flexibility to explore these new methods of content delivery, and to anticipate future collections and user needs, without requiring the expensive overhead of shifting to a completely new DAMS.

Rationale and Data Behind Recommendation/Testing Results

Fedora 4

As noted above, because Fedora, the repository underlying both Islandora (currently Fedora 3) and Hydra/Sufia (Fedora 4), functions as a component rather than a standalone DAMS, it was not separately tested. However, during testing and researching, the TF observed the following strengths and weaknesses:

Strengths of Fedora:

- Robust development community, under the umbrella of DuraSpace (with some possibility of integration with VIVO and DSpace)
- Forms the basis of several popular open-source DAMS, including Hydra and Islandora
- A flexible object model that is complementary to DSpace's more constrained model
- Implements the Linked Data Platform W3C recommendation with support for RDF expression
- Built-in durability functionality
- Implementation draws on local strengths with Java development

Weaknesses of Fedora:

- Requires a significant investment of developer time and support, potentially in addition to the contracting of support like the Data Curation Experts group. (An illustrative anecdote: The University of Houston digital asset management assessment group had initially recommended bringing up Fedora alongside their locally-developed user interface but moved away from the plan in favor of Hydra-in-a-Box, in part as a resources decision. Note that UH currently dedicates one developer and one metadata coordinator to the project at 90% FTE, with three other librarians working on the project at between 5-25%FTE, depending on stage. They anticipate several years at this staffing.)
- Requires community investment to gain fluency (including attending Fedora users group meetings and Fedora Camps)

Implementation: Interim Solutions

Given the local development requirements incurred in the TF's recommendation, we anticipate the need for interim digital asset management solutions as the organization builds up knowledge and constructs necessary components.

There are several options for meeting existing repository needs until a DAME can be realized. These include adding an out-of-the-box vendor product, bringing up a temporary Fedora

solution, or continuing with DSpace with additional enhancements and workarounds to meet current unsupported needs.

ResourceSpace and Nuxeo are both suitable out-of-the-box systems for an interim solution. They offer management of content that our current DSpace implementation does not support, such as streaming media and hierarchical metadata (Nuxeo-only) and complex relationships (Nuxeo-only), and could bridge the gap until a DAME is implemented. Even though they are out-of-the-box solutions, there will still be overhead in terms of time (setup and training of personnel) and costs (not insubstantial in the case of Nuxeo). As an interim solution, ResourceSpace, which is both more lightweight and more limited in its functionality, may be the more desirable out-of-the-box solution. It is available as open source software, and other campus units use or are interested in using ResourceSpace to manage content (primarily images). Knowledge gained from a ResourceSpace deployment could be used to achieve cross-system searching and retrieval of content from other campus entities using ResourceSpace by the DAME. Nuxeo would require time and considerable expense that would not be recovered.

Fedora architecture and its ability to handle complex metadata, relationships between objects, and versioning show considerable promise as a component of a DAME. The only quick solution for interim deployment of Fedora is Islandora, which relies on Drupal as its interface. Islandora functionality can be extended by adding Drupal modules; however, installing the modules and their dependencies is not always a straightforward process. In addition, this is a moment of change for Islandora. Islandora is currently transitioning from Fedora 3 to Fedora 4 and, in the process, is rewriting the Islandora stack and working towards interoperability with Hydra (via the Portland Common Data Model, which would provide a shared data model). Reliance on a vendor like DiscoveryGarden may help ease a short-term adoption of Islandora.

The Team identified several ways to extend DSpace functionality and allow it to serve as an interim solution. Video capability of our current DSpace could be extended by installing new video streaming tools developed at Virginia Tech. The need for completely private deposits, not visible to anyone, would be handled by use of private status in DSpace, depositing those items directly in Archivematica, or bringing up another instance of DSpace for dark storage.

Finally, there is clear demand for and interest in the deployment of an exhibit tool. The TF did not look closely at these options, though its overall impression of Spotlight, tested in conjunction with Blacklight/Hydra but deployable as a standalone product, was positive. The TF's Exhibit Sub-Group (members: Jeremy Brett, Cecilia Smith, Kevin O'Sullivan, Nancy Burford, Robin Hutchison), which has also looked at Scalar, Omeka, and other tools, is better positioned to make a recommendation.

Costs associated with recommendations

Fedora 4.

The TF did not perform a detailed cost analysis of a Fedora implementation based on desired features. Digital Initiatives would be able to generate this data upon request, through a resource-intensive estimation process. They would need to dedicate a considerable amount of time to mapping out what this investment would require, from bringing up Fedora to building tools and user interfaces. A rough estimate is that the project would require the equivalent of 2 FTE developers for the first year, or 1 FTE developer to conduct pilot testing and development. Given the complexity of the project, we also anticipate the need for 1 FTE librarian, with deep metadata expertise, to perform ongoing needs assessments, coordinate development, design tools, and implement policies and workflows. Additionally, we anticipate costs related to attendance of essential Fedora trainings and meetings, such as Fedora Camps, Open Repositories, etc.

Nuxeo.

Nuxeo provided us a quote that included a $100K annual fee comprising:

- Premium, platinum-level support
    - Software maintenance and hotfixes
    - Availability of senior architectural consulting services
    - 2 pre-production or production instances
    - 2 development instances
- Multi-user node in Nuxeo Studio (online application editing tool)
- Access to the Nuxeo Marketplace (to select ready-made plugins)

However, this fee does not include storage costs, which Nuxeo estimates to be ~$15K annually for 5TB / $25K for 10TB of hosting via Nuxeo-hosted Amazon Web Services. The University of California (UC) shared with us that their annual fee, currently $27K, has steadily increased. They were able to negotiate a discounted rate with Nuxeo because they locked in an earlier rate before prices skyrocketed. UC is currently contracted at a silver support level (for which the list price is $47K/year). While Nuxeo is technically an open source product that we could conceivably host and run on our own, UC has indicated that support is essential to effectively running the product. Their initial setup (0.5-1 year), required 25% FTE technical lead, 50% FTE of another developer (more focused on customizing the UI and creating user documentation), and 15-20% FTE metadata.

Additionally, for any new DAMS, we estimate that it will require a significant amount of coding and/or financial sponsorship of work to be done by Archivematica developers. Application development is required in order for Archivematica to deposit access copies of its content in a DAMS; specifically, to extract the content and metadata from the exported access copy (referred

to as the DIP – Dissemination Information Package) and repackage/ingest it in the designated DAMS.

Additional Considerations and Future Plans

The TF has emphasized the need for an ecosystem of services to meet the disparate and heterogeneous use cases of different types of digital assets. A case in point is audiovisual (AV) content. Users are accustomed to interfaces that allow the streaming of content in the browser without an explicit download or third-party application. Such facilities present technical challenges for small IT shops. In addition, metadata for AV content can be applied at a finer grain than for basic item records. Annotations can apply at the level of boxes or polygons in images, and at time ranges in video and audio. These metadata therefore require a hierarchical structure that is ill-served by key-value pair frameworks.

Linked Open Data (LOD) has garnered a lot of attention in the library community. This can provide a framework for deeply expressive metadata and discovery, but is not a silver bullet – just a better practice for representing information in a more interoperable way. The primary benefit of a LOD approach to managing information is the use of URIs to denote resources and relationships. Unique unambiguous identifiers are nothing new to the library community, but the URI is a construct that can bring new and legacy identifiers all together in the same namespace. Thesauri and name authority records provide the best identifiers for generally known entities, but as new or historic entities from the campus community are subjected to descriptions, the Libraries will need a local authority to mint identifiers. The PURL service can support this effort. Any entities that do not have URIs will not be addressable in the prospective LOD graph. It will require a significant and broad investment by various campus stakeholders to generate URIs for arbitrary entities that we want to have participate in our knowledge building exercise. The Libraries' mixed experience with VIVO deployment is an indication of the challenges in this area.

A RDF triplestore houses and serves the LOD graph, and Apache Jena is an open source product that could fulfill this role. The triplestore is a supplement to our existing information ecosystem that fills a new niche and could play a unifying role. Relational databases handle business and application logic, Solr indexes allow for quick searching, and triplestores allow for complex and highly interoperable expressions about resources with an expressiveness approaching natural language. The same information will be duplicated across these different storage systems for different purposes. A combination of these different approaches to information organization enables an unprecedented range of use cases and interfaces. However, these interfaces are largely unrealized at this point and any new features so enabled still require implementation work. This observation has been borne out by the TF's experience with the Hydra heads and their sometimes spotty feature sets, as well as by the lack of flashy interfaces to take full advantage of the VIVO triplestore.

A combination of all the different modes of data storage and robust APIs to access the data is the best way to facilitate and speed development of a UI like the UC's Calisphere to pull everything together for discovery. The Digital Aggieland Portal that Digital Initiatives has recently undertaken is an effort in this direction: the portal seeks to provide a unified interface to Texas A&M-related materials, including yearbooks and newspapers, currently scattered across different platforms, but will have to be developed concomitantly with the desired storage and APIs. It is worth noting as well that specialized storage solutions and APIs will better allow our data to be exported en masse for analysis.

Finally, as has been noted elsewhere, application development is required in order for Archivematica to deposit access copies of its content in a DAMS, specifically, to extract the content and metadata from the exported access copy (DIP) and repackage/ingest it in the designated DAMS.

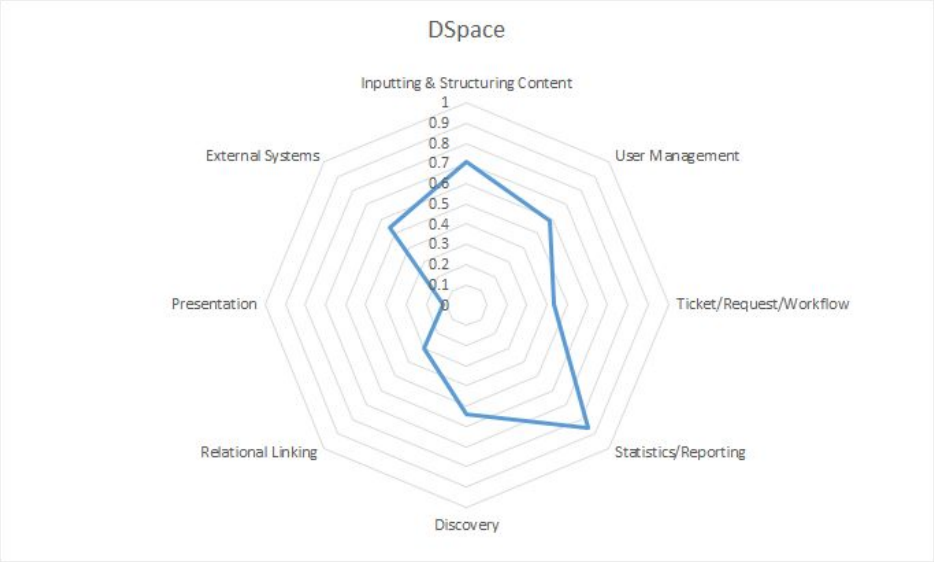
Findings from DAMS testing

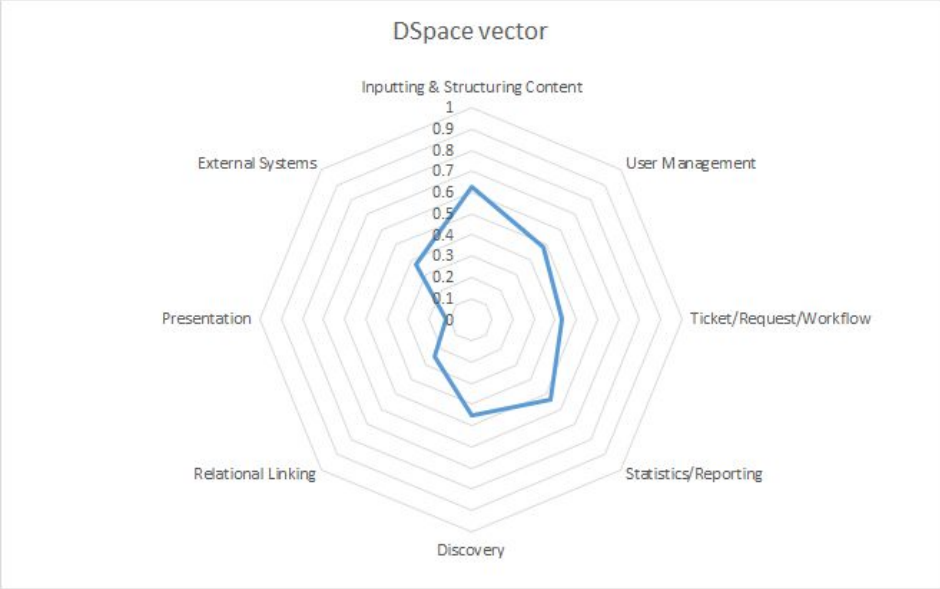
We selected four systems for close testing; additionally, we tested the newest version of DSpace. Here, we present a brief description of these systems, to include scores derived via rubrics, and the pros and cons of each. Every system tested boasted the following:

- a robust API
- broad adoption
- strong community support
- the ability to function as either a modular component in a DAME or a standalone DAMS (including a range of functionality and support for user management, display, indexing and discovery, built-in statistics, etc.)

DSpace

Version tested: 5.5
29 Configuration scores, totaling to 43.5; 23 Time-out scores, totaling to 23; and 143 feature scores, totaling to 249
Final score: 315.5
Vector space final score: 253





Strengths:

- Ability to authenticate and set granular permissions and restrictions
- Integration with Shibboleth (Texas A&M's user authentication system, which underpins NetID login), handle servers (for minting unique identifiers), and Archivematica
- Strong knowledge and investment based in the University Libraries
- Active international user and development communities, under the umbrella of DuraSpace

Weaknesses:

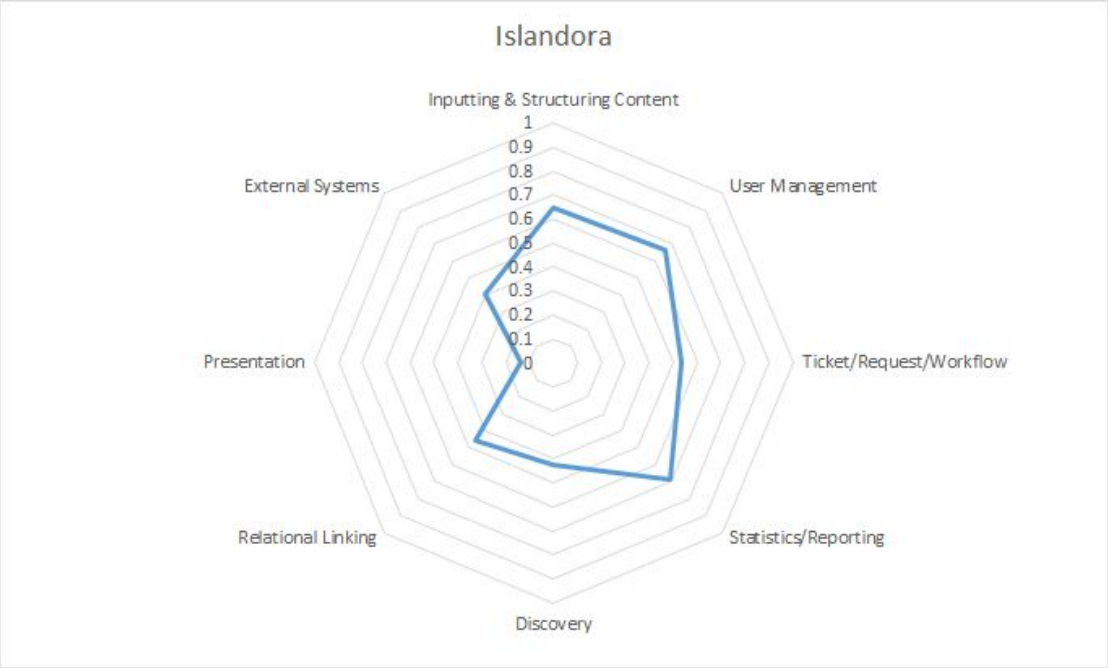
- Support for complex objects or relationships
- Support for hierarchical or structured metadata
- Support for linked data (though this is an evolving area)
- Metadata versioning
- Reporting and auditing

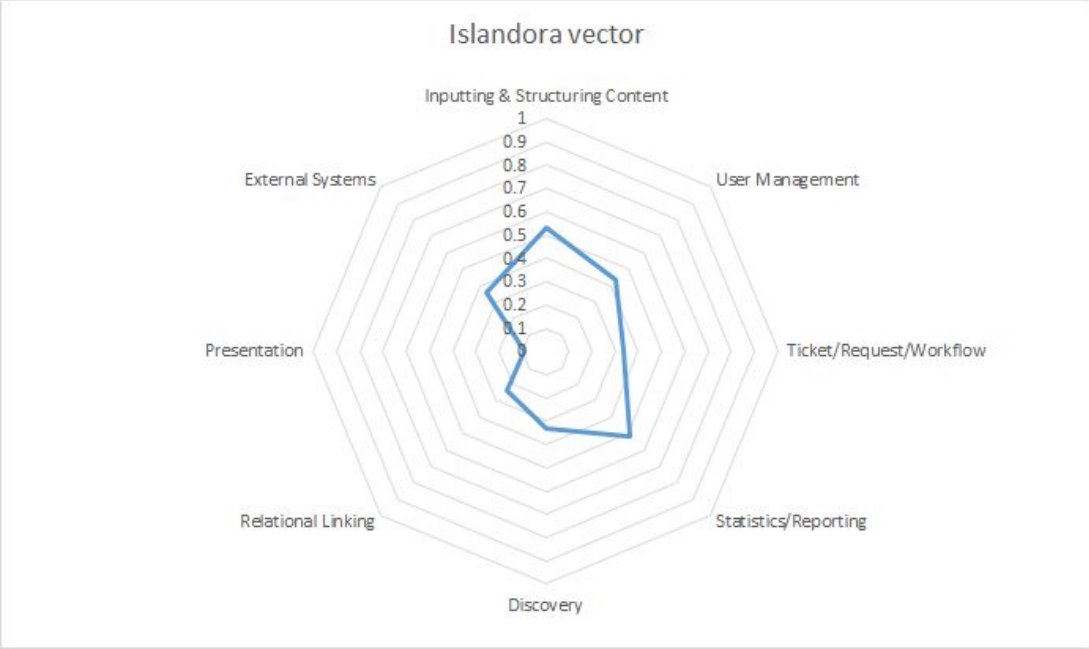
- Confusing documentation

Backed by a strong, international open source community, DSpace is continually evolving but has affirmed its function as an institutional repository platform rather than a fully-featured DAMS with robust AV capabilities. DSpace packages PostgreSQL and Oracle relational database storage layers with a bundle of business logic layers (including Lucene search) and application layers, including the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and a JSPUI user interface. Texas A&M has a wealth of experience with and long investment in this product. Given DSpace's lack of built-in functionality for non-print media collections and the limitations of the product's data model, we recommend diversifying by bringing up flexible DAME components, which could complement DSpace's core affordances.

Islandora

Version tested: 7.x.-1.6
59 Configuration scores, totaling to 88.5; 33 Time-out scores, totaling to 33; and 117 feature scores, totaling to 195
Final score: 316.5
Vector space final score: 216





Strengths:

- Tailored support for diverse file formats and complex objects
- Strong metadata versioning
- Support for configurable, hierarchical, and structured metadata
- Excellent documentation

Weaknesses:

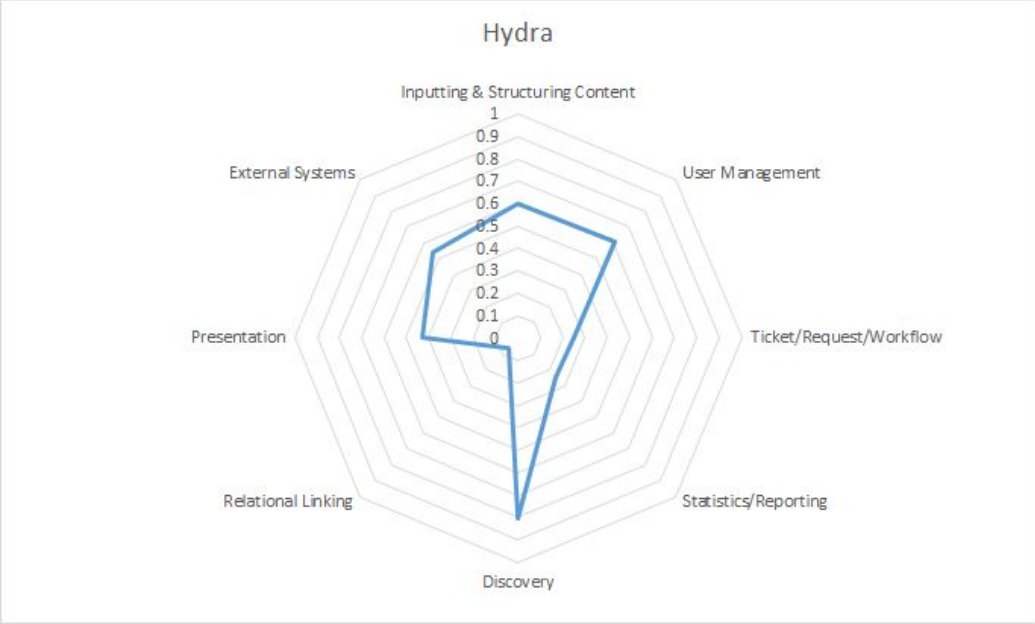
- Support for linked data (this is an evolving area)
- Ability to authenticate and set granular permissions
- Unintuitive interface
- Reliance on Drupal (could lead to future limitations)

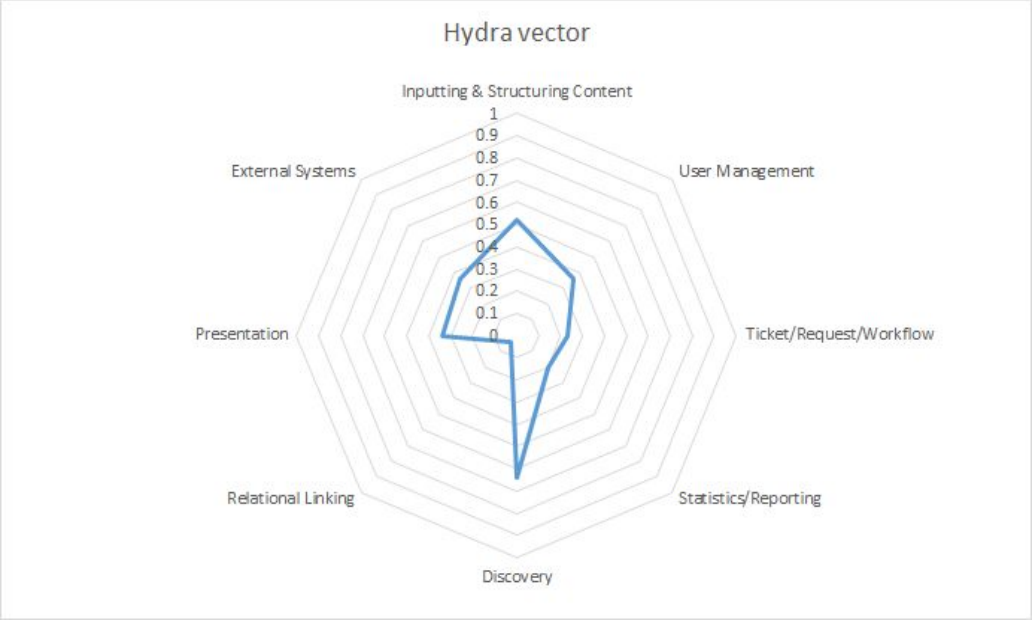
Islandora's open source community is somewhat smaller than that of DSpace, with some dependency on the DiscoveryGardens vendor. This product marries assets managed in a Fedora 3 repository (currently transitioning to Fedora 4) with Drupal modules for browsing; core hook and API functionality from the Drupal + Fedora Islandora "Core" enables interactions with "Solution Packs" for creating, viewing, and managing collections. Islandora elicited divisive responses from our testers: some were impressed by 'excellent and useful documentation,' full features, AV handling, and metadata versioning; others were put off by the 'unintuitive' interface and struggled even with simple upload. We suspected that Islandora could offer a 'tremendous amount of flexibility … once you overcome the sizable learning curve' but were stymied by the by-design inflexible object models.

Hydra/Sufia

Version tested: Sufia 6

46 Configuration scores, totaling to 69; 17 Time-out scores, totaling to 17; and 150 feature scores, totaling to 215
Final score: 301
Vector space final score: 226





Strengths:

- Support for diverse file formats
- Intuitive, well-designed user interface
- Functionality that we didn't see elsewhere, including: ability to easily transfer ownership of files, basic statistics in contributor profiles
- Integration with DropBox, Zotero, and other services
- Robust, engaged development community, under the umbrella of DuraSpace and the Hydra Project

- Potential for integration and growth with the anticipated development and deployment of Hydra-in-a-Box

Weaknesses:

- Support for hierarchical and structured metadata
- Support for complex objects (Sufia 6 does not allow for multi-part objects)
- Metadata versioning
- Ability to authenticate and set granular permissions
- Reporting

The TF chose to test the all-purpose Hydra head Sufia, currently available in version 6. Overall, we were taken aback by the limitations of the system, which did a few things very well and simply did not perform in a number of important areas. As one tester commented: "importing files was easy and intuitive, yet importing and associating metadata was seemingly impossible." Testers complimented the elegant user interface and ease of self-deposit, but were frustrated by the lack of batch deposit. Hydra is considered something of an open source darling, garnering impressive institutional commitments and grant funding for development; development and planning for the project appears to be well-staffed with dedicated, active volunteers. While some features have been implemented in local Sufia instances, Sufia 7 (currently under development and also expected to serve as the basis for Hydra-in-a-Box) promises to significantly expand functionality.

While testing focused on Sufia, we looked at Spotlight, Blacklight, and Fedora in the process. John Bondurant also looked at Avalon, the AV stack for Hydra developed at Indiana University and pronounced it as a strong standalone system for AV materials, handling management from ingest to preservation, with strong control over permissions and account set up.

ResourceSpace

Strengths

- Existing user community at Texas A&M
- Highly functional as a standalone content management system, used by many smaller institutions
- Particularly well-suited to handling images and other AV materials: functionality includes preview/streaming, support for alternative files, and inclusion of multiple versions/formats of files in a resource set
- Very easy to batch upload
- Simple, advanced, and geographic search options
- Metadata versioning
- Provenance and authority checks
- Statistics and reporting

Weaknesses

- Limitations of the simplified testing model/approach that we applied to Nuxeo and ResourceSpace make these systems difficult to fully assess
- Appears not to support handle or DOI minting
- Flat metadata / no hierarchical or structured metadata support / no linked data support

Originally developed by Oxfam in Great Britain, ResourceSpace is a web-based, open source LAMP-based (Linux, Apache, MySQL, and PHP) DAMS designed for ease of use and access. For relatively nominal annual fees, a company will provide training, support, and, if requested, hosting. Overall, testers were 'pleasantly surprised' by ResourceSpace, which handled AV materials nicely and worked well as an out-of-the-box solution. Testers commented on the ease of generating reports and statistics tracking usage, contributions, and other actions. The system offered support for user-contribution workflows, streaming audio and video, a geographic search functionality that we did not see built-in elsewhere, and easy-to-administer (if flat) metadata. The Texas A&M Foundation, which uses a free, open source version of ResourceSpace hosted on their own servers, offered a demo of their experience with the system (their use is tightly coupled with editing and preparing files using Adobe Lightroom). The Foundation has been happy with the tool, which provides a range of access options that they frequently deploy.

Nuxeo

Strengths

- Full-featured digital asset management system
- Includes features not seen in other systems, including editing capabilities, abundant automated metadata extraction, automatic creation of derivative files, and filters that manage assets through their stages of development and content types. Most intriguingly, there are highly customizable business logic models that enable configuration of object management over a lifecycle.
- Flexible, extensible metadata models
- Strong example from UC library system of the utility of the system, including excellent documentation and user guides
- Easier and faster to implement than open-source platforms based in Fedora

Weaknesses

- Vendor model that is necessary to enable real functionality in the system is expensive; these high costs are, according to the UC experience, likely to rise
- Though it appears that the system's functionality aligns with our needs, it was difficult to locate the TF's defined use cases in the documentation, which speaks to a larger, more corporate user
- Sparse cultural heritage user community
- Reporting difficulties handling complex objects

- Possibility of development to incorporate linked data and external vocabularies (but not yet fully featured)
- Limitations of the simplified testing model/approach that we applied only to Nuxeo and ResourceSpace make these systems difficult to fully assess

Nuxeo is an enterprise digital asset management system with little uptake in the cultural heritage or academic sectors. Accordingly, it stood out among the systems we tested for unique functionality that was a little harder to map to our needs. The University of California Digital Libraries (UCDL) group, who use Nuxeo as the administrative interface and DAMS to digital collections published through Calisphere, helped us understand the system's application in a peer institution. UCDL was generally quite positive about the system, indicating that they were able to go far, fast. While not quite out of the box, it took their group between 6 months and a year to bring up the system and perform some customizations, with most of that time going to prepare metadata and connect to a separate UI (they estimated a labor breakdown of 25% technical lead, 50% developer focused on UI and documentation, and 15-20% metadata).

Implications for Libraries and Campus

Responding to a survey, representatives from campus units, indicating both administrative and academic needs for digital asset management, tended to emphasize local storage and retrieval requirements. One of the advantages of a campus DAMS, however, is the ability to work across the full network of campus digital assets to locate and connect relevant resources. Deep links between campus collections can surface when we work at scale and focus on connections. The OneHealth Initiative, for example, given its multidisciplinary scope of human, animal, and environmental health, can fuel coalescing items/materials from across campus and the extension agencies. The drawing together, retrieval, and display of these items is strengthened by a strong linked data model with graphs, persistent URIs, etc. Examining and developing public policy issues of the Texas-Mexico border not related to immigration furnishes another example of a multidisciplinary academic area that would benefit from a shared asset system. Possible topics, building across expertise and holdings in the Bush School, PSEL, Cushing (including holdings of the System agencies studies), and the *Texas Colonias* program of the College of Architecture, include mass rail transit across the border, emissions studies from idling semi-truck engines waiting to cross the border inspection points, improved designs of those very same border points, and water rights and usage.

Campus Case Studies: The Technical Reference Center and Division of Marketing & Communication

Marketing & Communication

The Campus Division of Marketing & Communication (MarComm) has internal storage in place with a networked drive for still images and large capacity local arrays for high resolution, raw

video. They plan to keep those parts in place, as they are not interested in sharing all raw and untouched materials. A DAMS is needed to help MarComm store, organize, and tag the assets, then provide a convenient and user friendly gateway for internal usage, accessible to others within the division, such as graphic designers and web designers, as well as a gateway for external clients that enables search, preview, and download. Multiple resolution options should be made available for download. The group needs total control over which assets are made completely public and which are only internally accessible: the division will be offering select assets as opposed to their complete collection, especially in the video realm.

Technical Reference Center, College of Architecture

The Technical Reference Center (TRC) is an independent reference library supported by and housed within the College of Architecture. The TRC is charged with two primary responsibilities: providing teaching and learning materials to faculty and students and maintaining the College's historical archive. With the implementation of a campus-wide DAMS, digital collections produced within the College could be made searchable and accessible not only to College of Architecture personnel, but to a wider audience of scholars at Texas A&M University and beyond.

The TRC in particular has a wide range of needs that could be met by a flexible and capable DAMS. One major ongoing project seeks to digitize a collection of 150,000 35mm teaching slides. The slides depict art, architecture, construction, urban planning, and landscape architecture subjects. Each slide is labeled using the Simons and Tansey Slide Classification System, which describes the content in terms of chronological period, country, medium, style, origin/creator, subject, subject subdivision, title of work, and details. This descriptive metadata is currently being entered into a spreadsheet. Because there are a large number of images already digitized and a considerable amount of data already created, a DAMS that supports bulk uploads of both images and metadata is needed. Also, because some of the images are subject to certain licensing restrictions, flexible authorization control must be available.

The TRC is also responsible for maintaining the contents of the College's historical archive. A substantial amount of material in a variety of formats has been or will be digitized, including documents, photographs, and AV recordings. Thus far TRC staff has digitized 75+ videotapes, 300+ photographs of *Texas Colonias*, 700+ glass lantern slides, 400+ print items from the *Ernest Langford Collection*, and 4,400+ structural drawings from the Mosher Manufacturing Company (1903-1922). All of this material is currently residing locally on a redundant disk storage system. Like the teaching slide collection, all associated metadata is saved in spreadsheets and would need to be bulk uploaded.

To meet the current need for teaching images, the College of Architecture is running an instance of the Madison Digital Image Database (MDID) on a local server. The platform is now quite

dated and the volume of stored data has resulted in a frustratingly slow user experience. Furthermore, MDID is capable of handling only JPEG images and associated descriptive metadata; other file formats such as TIFF images, audio, video, PDF documents, etc. are not supported.

Beyond the needs of the TRC, there are other units in the College of Architecture that would likely benefit from access to a DAMS. The CRS Center for Leadership & Management in the Design & Construction Industry, for example, is a repository of the business archives, slide archives, oral history, publications, and architectural program library of the CRS architecture engineering and planning firm, and efforts to digitize this material are underway. Additionally, the College's Office of Communications maintains a large and growing repository of digital photographs and video recordings of College events, thus they have digital asset storage and retrieval needs similar to those of MarComm.

Libraries' Projects Waiting for DAMS/DAME Implementation

Preservation is capable of digitizing most formats including audio and video. The Director of Preservation has not been pursuing large scale digitization projects because of the difficulties associated with metadata and accessibility. Some projects are moving forward because the TF is closer to determining a DAMS and catalogers are working with Preservation to create metadata for both new projects and retrospective collections.

The Libraries' collections fall under the following categories for digitization:

1. Collections that have been digitized but whose files are sitting on a server with no accessibility
2. Collections that will be digitized this calendar year
3. Collections that we can and want to digitize, but we are waiting for the DAMS to be operational or near operational before starting

Audio and Video Collections

There are several collections within the Libraries' special collections and archives that are still under copyright; however, due to preservation concerns and accessibility needs these collections could be accessible online via a NETID login or through the Cushing Reading Room. These collections include, but are not limited to: (# refers to the category listed above)

- Cushing *SciFi and Fantasy Collection*; George R. R. Martin and Filk (#3)
- *Fred Cuny Collection* -- donor requested digitization and access (#1 and #2)
- Texas A&M University Archives; football films, oral histories, Forestry Service (#1 and #3)
- Area Studies; oral histories, *Yolanda Broyles-Gonzalez Collection* (#1 and #3)
- Literature; *William A. Owens Collection* (#1)
- *Clements Collection* -- Cushing request for digitization and access, over 1,000 items (#2)

Audio/Video Functionality

Beyond storing and streaming AV materials, the Libraries might develop the capability to mark particular places in an audio or video recording – a bookmark of sorts – for ease of access during playback or streaming. The mark or highlight would be visible on the progress bar for easy navigation and these "marks" could be aggregated for a scene selection submenu or search facet. For example, a system that will allow a search for every touchdown or interception in Aggie football videos.

Photographic Materials

Preservation has been asked to digitize and make accessible an estimated 33,000 images associated with the *Sports Information Archive*. Due to the scope and nature of this project, it has been determined that the DAMS would be the most appropriate location for these images to be managed. As a part of the DAME, exhibit software would be paramount for this collection to be used and accessed as intended by the Sports Media Relations department. (#2)

In addition to that large scale project, Preservation is finishing a digitization project for the Medical Sciences Library (MSL) in conjunction with the 100th Anniversary of the College of Veterinary Medicine (CVM). An estimated 6,000 images were digitized and will be placed into DSpace until the DAMS is available. (#1 and #2)

Cushing has several photographic collections that could be made available including the *Chatham Collection*. This collection has not been digitized but will be undergoing a condition assessment to determine preservation needs. The Director of Preservation is aware that much of this collection's negatives has moderate to significant acetate decay and digitization will be the only way to preserve and make these images accessible. This collection would need to be exhibited in the DAMS. (#3) Some photographic materials from Cushing have been digitized and reside in DSpace or in the Libraries' institutional Flickr account.

Additional photographic materials projects include but are not limited to:

- Low Speed Wind Tunnel historical images (#1 and #3)
- Bonfire slides -- donor requested digitization and access (#3)
- Graphic Services (#1)

Bound Collections

Several volumes from the Cushing collections have been digitized to improve access and preservation. Currently, none of these volumes are available online. Of significant note are the two large, wooden board Antiphonals and the thousands of images that have been generated for the *Cervantes Project*. The latter images are only available on the *Cervantes Project* website, which is considered a non-functional website and whose future is questionable. (Both #1)

More bound collections could be easily digitized and accessible for patrons and exhibits including one of the largest and thorough collections of historic veterinary and equine medicine volumes housed at the archives and special collections in the MSL. To make these materials accessible, the Libraries' needs a DAMS and exhibit software and Preservation has not pursued digitization of these materials because of a lack thereof .

Maps/Blueprints/Renderings

Large scale or oversize material is difficult to physical access and preserve. The Preservation Unit and the Libraries would like to rely more on digital access to protect these types of collections. Projects or collections including TAMU Archives facility renderings (#3), historic maps (#1 and #3), and campus maps which are only accessible in Mapping Historic Aggieland (#1 and #3).

Born-digital Materials

The changing medium in which documents, images, and AV materials are being produced calls for the need to handle the born-digital materials in archives and special collections. Texas A&M University almost exclusively produces records and content in the digital form. The University Archives are tasked with the preservation of the University's history which in turn means collecting born-digital materials.

Currently there are about 7.3 TB of born-digital materials identified from collections in Cushing. Several recent acquisitions of born-digital content such as *Basbanes*, SciFi manuscripts and fan videos, and digital images from A&M Health Science Center support the realization that the rate of born-digital content at Cushing is only going to increase. Current born-digital content includes digital documents, images, audio, and video from the SciFi, Hispanic, and Area Studies collections. Most, if not all, of this material is on floppy disks, CDs, and DVDs and will be preserved as forensic disk image files and placed in preservation storage (Archivematica). Some of this preserved digital material, depending upon conditions such as donor agreements and/or file format, will be included in the DAMS as part of the digital archiving workflow that is currently being developed.

Glossary

Application program interface (API): A set of routines, protocols, and tools for building software and applications. Web APIs are the defined interfaces through which interactions happen between an enterprise and applications that use its assets. An API approach is an architectural approach that revolves around providing programmable interfaces to a set of services to different applications serving different types of consumers.

Digital Asset Management: "... consists of management tasks and decisions surrounding the ingestion, annotation, cataloguing, storage, retrieval, and distribution" of image, multimedia, and text files.

Digital Asset Management Systems (DAMS): Typically, products that serve as the management layer in digital asset management ecosystems or digital libraries. These products differ in their approach to functions and range of associated capabilities.

Digital Asset Management Ecosystem (DAME): The term signals a philosophy of modular development, with attention to the ways systems need to work together to enable full functionality. The choice of the word ecosystem, as opposed to "system" (as with a DAMS) is explained by the DAME's emphasis on a distributed service architecture. This is an architecture in which the discrete roles of a DAMS are handled not by one application, but instead by a collection of applications, each one suited for the role it plays.

User Interface (UI): We use this term to refer to both the public and administrative interfaces to digital collections and systems. The public interface is likely to be focused on exhibiting, displaying, and extracting collections. The administrative interface deals with ingest, editing, setting permissions, etc.

Appendix 1: DAMS Task Force Assessment Process

The Digital Asset Management Task Force was approved by the University Libraries Executive Team on August 20, 2014 with representative volunteers for the TF selected by each Associate Dean. Shortly after the approval, a meeting with the Dean of the Libraries, Associate Dean for Information Resources, and the Associate Dean/Director of Cushing Memorial Library and Archives was called to determine the scope and objectives of the project.

The group began by conducting literature reviews and environment scans to (1) investigate digital asset systems and (2) DAMS need/selection assessment processes at a number of peer institutions. Consultations with and documentation provided by the University of Utah, University of Houston, and Penn State University were particularly helpful, as were reports out of Michigan State University, the University of California, and the National Institutes of Health.

To ensure that all appropriate and available digital asset management systems were reviewed, the TF researched and created a list of 26 possible systems. Members of the TF, which switched from biweekly, 1-hour meetings to weekly, 1.5-hour meetings in March 2015, reviewed each of these systems in depth to identify the license type (open source or proprietary), what organization or institution developed and managed it, institutions that used the systems, and additional anecdotal information from articles or case studies. In some cases, TF members spoke directly with users of these systems or participated in product demos. Members of Digital Initiatives (DI) reviewed the systems further to determine if they would be compatible to other programs and software used throughout the Libraries (ex. Java based). Please *see Appendix 2 for a list of the initial 26 systems.*

It was during this initial review of the 26 possible systems that the TF was contacted by others on campus (including MarComm, The Office of Former Students, and May Business School, to name a few) who had either brought up a DAMS or were looking into bringing one up. These other entities wondered if one campus-wide system could be brought up for everyone's needs and if the Libraries should lead this endeavor due to the inherent large-scale needs of the Libraries. This concept was new to the TF and required additional planning, scope development, and a survey to determine the needs of the campus. The TF designed and ran a survey February-April 2015 to gather data on campus use cases. *Please see Appendix 3 for the campus-wide DAMS needs survey.*

With the new needs and criteria for a possible campus-wide system now defined, the TF continued to review and pare down the list of 26 systems to five final systems. During this time, the TF also created multiple DAMS use cases gathered from the campus survey as well as Libraries-based needs, which formed the basis of work to develop a grading rubric.

The TF narrowed the viable options down to five systems, four of which were open source and one was an open source platform (Nuxeo) but with an accompanying online services subscription designed to deliver initial and ongoing operational support and maintenance with a customized suite of modular tools and Cloud storage, all at a substantial cost. DSpace is the Libraries' current access and repository system and was one of the systems selected, in large part because of the high level of expertise and experience had by DI. The other three systems selected for testing were Islandora, Hydra (with Sufia and Blacklight), and ResourceSpace. In November 2015, two campus testers joined the task force: Ian Muise (College of Architecture) and Brian Wright (MarComm).

The TF decided to pilot each system individually and sequentially with a common rubric and using multiple pre-determined sets of sample content containing various types of files, to include AV. The rubric developed by the TF consisted of individual tasks or functions grouped into the following eight sections, each with additional sub-sections:

1. Inputting and Structuring Content
2. User Management
3. Ticket, Request, and Workflow
4. Statistics and Reporting
5. Discovery
6. Relational Linking
7. Presentation
8. External Systems

Each TF member was assigned multiple subsections from the list above to test sets of discrete tasks across all systems, and each subsection was tested by two or three TF members. All TF members assigned to a subsection then collectively gave each specific test item a final rating. Members used a grading scale between 0 (low score) - 3 (high score) to mark how easily tasks on the rubric could be completed. In addition, the letter "C" was used in conjunction with the numbers to clarify if something could be configured or further developed locally or by the community, according to systems documentation or other research. The letter "T" was used for "timed out" if a member spent more than 20 minutes looking for the solution and could not find it. Notes were also gathered in the spreadsheet to help members testing the same task to communicate with each other and to keep track of research to help determine if a task could be configured. *Please see Appendix 4 for the rubric template and Rationale and Data Behind Recommendation /Testing Results for summary of testing results.*

The systems were developed, deployed, and tested, generally in one-month intervals, in the following order: DSpace, Islandora, Hydra (+Sufia, Blacklight/Spotlight), Nuxeo, and Resource Space. For the first three pilots, DI developed a sandbox/test environment for TF members, providing them with accounts/logins and technical support, when necessary. However, because

of technical problems with Islandora, many tests were performed on a hosted sandbox rather than our local testbed. TF members tested all assigned sets of tasks using the rubric and with the pre-determined sets of sample content. We were unable to deploy a fully viable version of the fourth system, Nuxeo, because of the customization of the "Nuxeo Studio," the costly web-based subscription service and Cloud storage, which could only be configured and demonstrated by Nuxeo developers. Subsequently, Sarah Potvin coordinated a professional, customized demo of the system with Nuxeo representatives on behalf of the TF. The Webex demo went well; however, because the TF could not evaluate the system using the same rubric and content as the previous three, and because of increasing time pressure to complete testing, a modified and condensed rubric was created to score the Nuxeo system. Nuxeo assessment further benefitted from an informative conference call that the TF held with a UCDL group that had recommended the deployment of Nuxeo and brought up the system. A shortened rubric was also used to score the final system, ResourceSpace. Additionally, a demonstration was offered by an employee of the Texas A&M Foundation who oversees their ResourceSpace instance.

Having three systems graded using the "long" rubric and the other two using the "short" rubric proved problematic. Of course, the short rubric scores could not be converted to a long rubric equivalent, which captured greater granularity and depth. Effective and uniform scoring could only be done by converting the long rubrics to short ones. The three long rubric scores were thus converted using two different approaches, allowing all five rubrics to be compared against one another and the data analyzed. *Please see Appendix 5 for the simplified rubric template.*

The DI members of the TF graciously accepted the task of analyzing the data and producing charts to represent systems comparisons. Utilizing the data figures, the TF identified the strengths and weaknesses of the various systems and all members were granted the opportunity to express their impressions and opinions about which system(s) would best suit the Libraries'/University's DAMS needs. The group discussed the findings and opinions and collectively decided upon a long-term recommendation for a modular or multi-component Digital Asset Management Environment (DAME), with potential interim solutions for more immediate DAMS needs in the short term.

Appendix 2: Systems Considered

- Mnesys
- ContentDM
- Greenstone
- Omeka [as presentation layer]
- Fedora Hydra
- Fedora Islandora
- Nuxeo
- ResourceSpace
- XTF
- MDID
- SharedShelf
- KORA
- Luna (added 3/30/2015)
- Cumulus (added 3/23/2015)
- DSpace
- EPrints
- TACTIC
- VITAL
- Invenio
- DigiTool
- Yoolib
- DSpace
- ORI-OAI
- ArchivalWare
- DAITSS
- Keystone DLS

Appendix 3: Campus Needs Survey Questions and Results

Survey Questions

Texas A&M University Libraries: Campus Digital Asset Management Needs Assessment

Welcome! This needs assessment is aimed at helping the TAMU Libraries Digital Asset Management Systems (DAMS) Task Force understand campus digital content needs, the first step towards forming a campus-wide initiative to preserve and maintain digital assets.

> **Digital assets**: Computer files stored on digital devices for subsequent access and use. Common digital assets include images, texts, presentations, data sets, and audio/video recordings. For the purpose of this survey, 'digital assets' do not include items that are produced in the course of normal business operations whose retention is dictated by records management policies, or items that are considered disposable.
> **Digital Asset Management System (DAMS):** An intertwined structure incorporating both software and hardware and/or other services in order to manage, store, ingest, organize, and retrieve digital assets.

Please provide the following contact information. The Libraries DAMS Task Force may contact you with follow-up questions.

Does your unit currently have storage capacity, or do you rely on others for storage? Please elaborate:

For the remainder of this survey, "unit" will reference your department, unit, etc.-- please answer questions based on that level of organization.

Digital Content

1. What types of digital materials does your unit produce that need to be stored, accessed, or preserved? This category should <u>not</u> include items that are produced in the course of normal business operations whose retention is dictated by records management policies, or items that are considered disposable. Please check all that apply.
   - ❑ Word Processed Documents
   - ❑ Imaging—Paper Documents
   - ❑ Imaging—Photos
   - ❑ Imaging—Non-Photos (e.g., maps, drawings)

- ❑ Digital Photos
- ❑ CAD Drawings
- ❑ Digital Graphical Images (e.g., digital maps, digital drawings)
- ❑ Audio
- ❑ Video
- ❑ Spreadsheets
- ❑ Databases
- ❑ Presentations
- ❑ Web Pages (if possible, please specify)
- ❑ Data Sets
- ❑ Other (please describe/list):

2. Approximately how much digital content does your unit maintain?
   - ❑ < 1 TB
   - ❑ 1-5 TB
   - ❑ 5-10 TB
   - ❑ > 10 TB [please approximate:]

3. How is your digital content stored? Please check all that apply.
   - ❑ Hard drive
   - ❑ Removable Magnetic Media (e.g., floppy discs, Zip discs)
   - ❑ Optical Media (CD/DVD)
   - ❑ Digital Tape
   - ❑ Solid State (e.g., flash drive)
   - ❑ Networked Storage / Server Space
   - ❑ Cloud Storage
   - ❑ Other (please specify):

File Formats

4. What file formats are created and/or maintained by your unit? Please check all that apply.
   - ❑ Text (RTF, ASCII)
   - ❑ PDF
   - ❑ HTML
   - ❑ XML
   - ❑ TIFF
   - ❑ JPEG
   - ❑ RAW
   - ❑ WAV
   - ❑ MS Word
   - ❑ MS Powerpoint
   - ❑ MS Excel
   - ❑ MS Access
   - ❑ MS Publisher
   - ❑ MS Visio
   - ❑ Other (please specify)

Technological Infrastructure

5. What is your unit's current storage capacity?

6. Does your unit plan to expand capacity? Please elaborate:

7. Do you use any content management, specialized software, or digital asset management systems to manage digital files? (e.g., SharePoint, DSpace, ResourceSpace, FileMaker Pro, etc.)
   - ❑ Yes
   - ❑ No

8. If so, which digital asset management system(s) are used?

Confidentiality Issues

9. Is any of your digital content of a confidential or sensitive nature? Please describe:

Access and Use

10. What would be your intended use for a new, shared digital asset management system? Check all that apply.
    - ❑ Preservation (including archiving, version control, file format conversion, etc.) [add definition as pop-up: Maintain digital assets for as long as required and in a form that is authentic and accessible.]
    - ❑ Organized online visual exhibits
    - ❑ Openly post materials online with descriptive information and/or indexing
    - ❑ Teaching (including streaming)
    - ❑ Creating online collections
    - ❑ Linked data/collaborative projects
    - ❑ Storage (not for preservation) [add definition as pop-up: Safely retain digital assets for future retrieval.]
    - ❑ Back-up (not for preservation) [add definition as pop-up: Create and store additional copies of digital assets in case original files are lost or destroyed.]
    - ❑ Other?

11. What levels of access do you require for your materials in the new digital asset management system? Check all that apply.
    - ❑ Openly available
    - ❑ Available for editing (crowdsourcing, tagging)
    - ❑ Available for download
    - ❑ Read only
    - ❑ Restricted access to select individuals or audience
    - ❑ Temporarily restricted (under embargo)
    - ❑ Not accessible
    - ❑ Other?

Libraries Deposit

12. Is your unit interested in contributing digital assets to the University Archives, for public access and/or long-term preservation?

Thank you!

Thank you for participating in this questionnaire. If you have any questions about the TAMU Libraries DAMS Task Force, please us at: DigitalAssetManagement@library.tamu.edu

Summary of survey results

Who filled out the survey?

1. Director, Web Development and Director of Visual Media (Marketing & Communications)
2. Interim Director of Marketing, Communications and Alumni Relations and Graphic Designer/Photographer (Mays Business School)
3. Graphic Designer/Photographer (The Association of Former Students)
4. Web and Information Designer (Physics)
5. Geospatial Librarian and Marketing Manager (University Libraries)
6. Communications Coordinator (Division of Student Affairs-University Center & Special Events)
7. Creative Manager (Division of Research)
8. Visual Resources Curator (Technical Reference Center, College of Architecture)
9. Director of Communications (College of Liberal Arts)
10. Director of Public Relations (Office of the Provost - Public Partnership & Outreach)
11. Communications Coordinator (Admissions)
12. Associate Director (Student Activities)
13. Senior IT Professional II (College of Geosciences)
14. Interim Exec. Director of Marketing & Creative Director (Marketing & Communications)
15. Marketing Manager (Texas A&M Foundation)
16. Graphic Designer/Photographer (Communications - College of Education and Human Development)
17. Web Developer (TEEX - Communications and Production)
18. Network Systems Manager (College of Veterinary Medicine)
19. Web Communications Specialist (Transportation Services)
20. Marketing & Communications Manager (Vice President for Student Affairs)

So:

1. marketing & communication professionals in:
    a. Marketing & Communications

b. Mays Business School
c. The Association of Former Students
d. University Libraries
e. Division of Student Affairs-University Center & Special Events
f. Division of Research
g. College of Liberal Arts
h. Office of the Provost - Public Partnership & Outreach
i. Admissions
j. Texas A&M Foundation
k. College of Education and Human Development
l. TEEX
m. Transportation Services
n. Vice President for Student Affairs

2. IT or academic professionals in:
    a. Physics
    b. University Libraries
    c. Technical Reference Center, College of Architecture
    d. College of Geosciences
    e. College of Veterinary Medicine
3. unclear
    a. Associate Director, Student Activities

Material types (23 responses):
Digital Photos (19)
Imaging—Photos (19)
Video (15)
Digital Graphical Images (e.g., digital maps, digital drawings) (14)
Audio (12)
Imaging—Non-Photos (e.g., maps, drawings) (12)
Presentations (10)
Word Processed Documents (9)
Imaging—Paper Documents (6)
Spreadsheets (5)
Databases (4)
Web Pages (4)
CAD Drawings (3)
Data Sets (1)

How much digital content?
<1TB (5)
1-5TB (8)
5-10TB (3)

>10TB (5) [estimates: unknown (3), 22TB, 12TB]

Storage type (22 responses):
Networked storage/server space (22)
Hard drive (13)
Optical Media (CD/DVD) (8)
Removable Magnetic Media (e.g., floppy discs, Zip discs)(2)
Solid State (e.g., flash drive) (2)
Cloud storage (2)
Digital Tape (1)
Other (please specify): External hard drive (1); ResourceSpace (1); Flickr (Libraries);

DAMS systems:
None specified (13+ 2 incomplete)
ResourceSpace (6)
FileMaker Pro database for archived tapes (not used for active file management) (2-- but same group)
Picasa with Adobe Bridge (1)
MDID2 (1)
Cloud (1)
Flickr (1)

Appendix 4: Rubric Template

To see a full rubric, visit
https://docs.google.com/spreadsheets/d/1E8EVi_W2ZQbtqbqxIiQdyxmXK8HnkuhAp1khpvClshU/edit?usp=sharing

Tests were graded on the following scale:
0-failure/impossible
1-possible/difficult
2-acceptable
3-excellent/easy as pie
T-timed out after 15 minutes
C-configurable (please include link to documentation)

Below are the tasks included in the testing rubric:

Inputting and Structuring Content

1. 1A. Set A. Albatross. Unnumbered/unordered set of of image files with homogeneous metadata. Includes at least one item with a license. NOTE: If this test cannot be successfully completed, no need to proceed-- this DAMS does not meet basic criteria.
2. Upload each item individually with a description
3. Search for the item
4. View each item
5. Preview each item without downloading
6. Update the file for one item.
7. Does (and if so how) the system indicate that the file has been updated?
8. Change the file for one item.
9. Does (and if so how) the system indicate that the file has been changed?
10. Update the record for one item.
11. Does (and if so how) the system indicate that the record has been updated?
12. Does the system retain previous versions of metadata?
13. Delete one item
14. Create a collection named Albatross
15. Create a descriptive landing page for Albatross
16. Upload the files in batch into this collection
17. Search across the collection using canned search terms
18. Create a second collection, Abalone, for a batch of resources
19. Import files into the Abalone collection
20. Order the files
21. Re-order the files

22. Export the Abalone collection
23. 1B. Inputting and Structuring Content. Set B. Bowerbird. Unnumbered/unordered set of files, heterogenous file formats.
24. Install a controlled vocabulary using a URL with a dynamic controlled term list-- use Geonames or VIAF
25. Install the (provided in Set B folder as CSV) list of controlled subject terms in the system / relate the list to the system
26. Change the static list of controlled subject terms
27. Create a form or rule at the collection level so that every item in the collection is assigned metadata indicating that the item is part of the Bowerbird Collection, Texas A&M University
28. Upload each item individually with metadata
29. Test: is it possible to upload the items with non-standard metadata? what processes are necessary for these items to be included? (note that one item is missing a filename, another is missing author/date/title)
30. Test: does the system prevent you from uploading the item without metadata? is it possible to designate required metadata fields? what processes are necessary for this item to be included? is it possible to search for or group items with incomplete metadata?
31. Change the order of the author names for an individual item
32. Upload a license as an attachment to an item
33. Assign additional subject terms using dynamic list
34. Assign URI as subject term using dynamic list
35. Assign additional subject terms using local controlled list
36. Affiliate an author name with an ORCiD, pulling dynamically from ORCiD registry (you can use the item by "Sarah Potvin" as a test)
37. Add an author name and the author's predetermined ORCiD to an item
38. View each item
39. Preview each item without downloading
40. Test URI included in item record: does it link out? [y/n]
41. Identify an item that has a specialized attached license
42. View the license attached to item
43. Are licenses automatically generated in single-item submission? [y/n]
44. Update the file for one item
45. Change the file for one item
46. Change the license attached to an item
47. Update the record for one item
48. Delete one item
49. Delete one license
50. Update the records for the collection -- add "Jr" to all contributor names
51. Create a collection for a batch of resources, named Bowerbird (B)

52. 1C. Inputting and Structuring Content. Set C. Cardinal. Numbered/ordered series of files, homogeneous file formats, licenses.
53. Create a collection for a batch of resources, named Cardinal (C)
54. Import files as a batch into the DAMS
55. Attach a license to the entire collection [this may be part of batch import or a separate step]
56. Change the collection license
57. Change the license for a particular item in the collection
58. Overarching thoughts on licenses in this system for batch load and collections
59. How does the system indicate, through reporting or metadata, that this item has changed since publication?
60. Order the files within an item
61. Re-order the files within an item
62. Order the items
63. Re-order the items
64. Using an Item from set B, cite an Item from set C [ideally, a link; possibly, a metadata field]
65. Confirm that the above citation relationship is evident on the cited item in set C
66. Export the collection
67. Is it possible to export at Web UI? Command line? Via REST? If so, perform one set of the following tests using each export method.
68. Export Dublin Core, METS, and MODS metadata for an item.
69. Using WebUI, export Dublin Core
70. Using WebUI, export METS
71. Using WebUI, export MODS
72. Using command line, export Dublin Core
73. Using command line, export METS
74. Using command line, export MODS
75. Using REST, export Dublin Core
76. Using REST, export METS
77. Using REST, export MODS
78. Export metadata for (a) an item; (b) a collection; (c) user-designated set of items as (1) JSON, (2) CSV, (3) RDF [9 tests total of feasibility].
79. export item as JSON (can be a system-specific format)
80. export item as CSV (can be a system-specific format)
81. export item as RDF(XML, TTL, or other well-supported serialization)
82. export collection as JSON (can be a system specific format)
83. export collection as CSV (can be a system-specific format)
84. export collection as RDF (XML, TTL, or other well-supported serialization)
85. export user-designated set of items as JSON (can be a system specific format)
86. export user-designated set of items as CSV (can be a system-specific format)

87. export user-designated set of items as RDF (XML, TTL, or other well-supported serialization)
88. Change the JSON, CSV, RDF crosswalks [3 tests: is it possible?]
89. Change the crosswalk, i.e. exported format for JSON
90. Change the crosswalk, i.e. exported format for CSV
91. Change the crosswalk, i.e. exported format for RDF
92. More qualitative assessment: examine the exported metadata and evaluate.

User Management

1. In the system, attempt to create and define a visitor, contributor, and administrative role, independent of assignment to either a specific user, or set.
2. Assign each role to a user.
3. View a list of all users and their roles, permissions, and usernames/IDs
4. Attempt to limit the read access to Albatross collection to administrator, Bowerbird collection to contributor, and Cardinal collection to visitor.
5. For visitors to the Albatross collection, does a message display indicating access level requirements to read?
6. Choose an item in the Cardinal collection. Restrict read access at the item/record level.
7. Choose an item in the Cardinal collection. Restrict read access at the file level.
8. Choose an item in the Cardinal collection. Restrict create access at the item/record level.
9. Attempt to limit the write access to Albatross collection to administrator, Bowerbird collection to contributor, and Cardinal collection to visitor.
10. Attempt to limit the delete access to Albatross collection to administrator, Bowerbird collection to contributor, and Cardinal collection to visitor [depending on how system is set up, delete access may be covered by write/edit access].
11. Attempt to grant access to a specific user with the role of contributor to Albatross collection.
12. Attempt to grant access to an item in the Dodo collection to visitor.
13. Set an account for the Albatross Collection to limit or allow batch file upload.
14. Set an account for the Albatross Collection to limit or allow batch delete.
15. Set an account for the Albatross Collection to limit or allow batch creation.
16. Find the item in the Cardinal Collection with a special license
17. Restrict read access to that item
18. 2B. User Management. Authentication Test.
19. Register a user and login using username/email and user defined password (Note: all DAMS should pass this test, but testing this will allow us to grade a DAMS on the relative ease of the login process)
20. View user profile (Note: this test should determine if the DAMS in question exposes the user as a meta presence)
21. Edit user profile (Note: this test should determine the degree to which the DAMS allows for the customization of the user's meta presence)

22. View your submissions from your profile page (Note: this test seeks to determine the utility of the users meta presence to other users)
23. Reset your password
24. 2C. User Management. Research Test.
25. Can the system log in with LDAP?
26. Can the system log in with Shibboleth?
27. Using attributes in Shibboleth, can the system automatically assign permissions for content based on rules?
28. Can permissions be limited by IP?
29. Is two-factor authentication needed?

Ticket/Request/Workflow

1. 3A. Ticketing/Request.
2. As a visitor, create a request for read access to the Albatross collection, once it has been limited to administrators.
3. As a visitor, report a bug or request a feature
4. As an administrator respond to the above report/request
5. 3B. Workflows.
6. As an administrator, configure the workflow for the DAMS to include a new metadata field in single-item submission
7. As an administrator, make another new metadata field [dc.subject.madeup] required for the Albatross Collection
8. As an administrator, change the label of this field as it appears on the input form
9. As an administrator, prompt a specialized metadata form to appear for AV materials in Set D // Dodo Collection
10. As an administrator, configure a special license to be made available as an option for items submitted to the Cardinal Collection
11. As the collection administrator of the Bower collection, require that all single item submissions be approved.
12. 3C. Research test.
13. Could the DAMS integrate into existing DI tracking software, i.e. Sugar?
14. Does the system allow users without permission to read items to request access?
15. Does the system allow users to provide feedback? Is it possible to set the email address to which these requests are forwarded?

Statistics/Reporting

1. 4. Statistics/Reporting. Statistics.
2. As a user, view statistics at the item level in the Albatross Collection
3. As a user, export statistics at the item level in the Albatross Collection
4. As a user, view statistics for the Albatross Collection. Do they include all of the items within the collection?

5. As a user, export statistics for the Albatross Collection
6. As an administrator, view and export statistics for the entire DAMS: Number of files and items
7. As an administrator, view and export statistics for the entire DAMS: Total file size of DAMS
8. As an administrator, view and export statistics for the entire DAMS: Formats of items in DAMS
9. As an administrator, view and export statistics for the entire DAMS: Date range of when items were uploaded into DAMS
10. As an administrator, view and export statistics for the entire DAMS: Number of times system/collections accessed
11. As an administrator, view and export statistics for the entire DAMS: Number of times items were viewed or downloaded
12. As an administrator, track and report the crud functions of an individual user
13. As an administrator, add a metadata value for every item created by that user
14. As an administrator, delete all records created by that user
15. As an administrator, track and report the crud functions of all users
16. As an administrator, track and report the crud functions of the contributor user group
17. As an administrator, track and report the crud functions performed in the Bowerbird Collection
18. As an administrator, track and report the actions of a particular user in the Bowerbird Collection
19. 4B. Statistics/Reporting. Research tests.
20. Is there a reporting feature?
21. Does the system track and allow reports to be created to show changes to metadata records made after a specific date or since last time record was exported?
22. Can custom reports be created?
23. If custom reports cannot be created, is it possible to query, via SQL, for instance, the sought-after data?
24. How is the report formatted / structured and disseminated?
25. How to request reports?
26. Limited to admin or certain users?
27. Web interface?

Discovery

1. 5A. Discovery. Search tests.
2. Using a text based search, discover an item from set Dodo using the items title.
3. Using a text based search, discover an item from set Dodo using the items author.
4. Using a text based search, discover a range of items from set Dodo using the publication dates.
5. Using a text based search, discover an item from set Dodo using the items subject.

6. Using a faceted search, discover all items of a specific file format (image, video, text, audio, etc.)
7. Using a faceted search, discover all items of a specific file type (mp3, tiff, jpeg, pdf, word, etc.)
8. Using a faceted search, discover all items of a specific subject (Football, Dog, Reveille, Texas Revolution, World War I, Correspondence, etc.)
9. Browse for items by date and date range
10. Browse for items based on creator
11. Add an arbitrary tag to an item in the Dodo set.
12. Discover the tagged item using a text based search.
13. Discover the tagged item using a faceted search.
14. 5B. Discovery. Research tests.
15. Does the system use solr indexing?
16. Can the system use a triple store?
17. Does the system index tuning through the ui?
18. Does the system allow for custom index solutions?

Relational Linking

1. Create an item for set D as follows: Take the files that constitute the book and add metadata to designate different parts of the book's hierarchical item structure as described in the provided XML metadata.
2. a) Designate specific pages as having annotations.
3. b) Designate page ranges as being sectional structures like chapters and sections.
4. c) Designate the exterior views (cover, back, etc.).
5. Group sets C and D into one parent collection. Are you able to navigate in the UI to the collection and see C and D as members? Do C and D link back the new parent collection?
6. Create a minimal item in D, i.e. a metadata record (bitstreams optional). Can you link a pre-existing item in D to the record to say the pre-existing item stands in each one of these relations to the new record: part-of, edition of, cites, referenced-by.
7. Do the links designate their targets as URIs?
8. Create a license for the parent collection. Can you find it being applied to collections C and D? Can you find it being applied to individual items in C and D?
9. Create an additional metadata field value on the parent collection that you would like applied to C and D and their contents. See if it is applied to the records for C, D, and their members.
10. Similarly for an authorization policy at the parent collection level.

Presentation

1. 7A. Presentation. Media players.
2. Does the system have native streaming video or audio functionality?

3. Does it integrate with systems that will enable this?
4. Does it integrate with image servers?
5. Does it integrate with image servers with IIIF compatibility?
6. 7B. Presentation. Exhibits.
7. Does the system have native exhibit capabilities?
8. Create a new exhibit, with a title and landing page.
9. Is it possible to select particular fonts for the text, creative custom headers or change the branding?
10. Pull items from Albatross, Bowerbird, Cardinal, and Dodo into the exhibit.
11. Does the system pull the metadata for these items in addition to the files themselves?
12. Assign new terms or customize the metadata that displays for these items in the context of the exhibit.
13. Design a browse feature that allows you to group items by existing metadata (type, title, creator)
14. Add metadata for the holding institutions to each item. Is it possible to browse by this new metadata?
15. Edit the title of the exhibit.
16. Edit the metadata for an item in the exhibit.
17. Publish the exhibit.
18. Assign a DOI to the exhibit.
19. Export the entire exhibit.
20. Delete the entire exhibit.
21. Research: what are the reporting functions for the exhibit?
22. Research: is there workflow functionality built in?
23. Research: how does the exhibit maintain a connection back to the original files/items in the DAMS?
24. Does the system integrate well with:
25. (1) Omeka?
26. (2) Scalar?
27. (3) Spotlight?
28. (4) GeoBlacklight?
29. (5) Blacklight?

External Systems

1. 8A. External Systems. API.
2. Confirm that there is a REST API
3. Confirm that all crud action are exposed through the REST API
4. Can these actions be authorized or authenticated?
5. Anything else you want to mention about the API?

6. 8B. External Systems. ORCiD. Note: Because of the emergent nature of ORCiD, these tests might be conducted as research into growing partnerships between DAMS and ORCiD.
7. Can you add an ORCiD as metadata natively? [tested in part 1]
8. Is it possible to add a field in the metadata registry to denote the ORCiD?
9. When adding an item by an author with a predefined ORCiD, can you associate the ORCiD with the name?
10. Can you enter an author name and dynamically pull an ORCiD from the central registry?
11. Can you pull a preferred author name from ORCiD?
12. 8B. External Systems. Miscellaneous.
13. External vocabularies
14. Integration with external controlled vocabularies (GeoNames, VIAF). Basic test is integrated into controlled vocabulary tests in part 1. Research test will need to look more specifically at whether and how DAMS can interact with these lists.
15. Archivematica. System can upload DIPs produced by Archivematica: identifies package components and structure, accepts derivative file formats, reads and transfers metadata records/files, etc.
16. Archivematica. System maintains correlation between ingested DIPs (from Archivematica) and content/metadata record(s) in DAMS
17. Archivematica. Can export metadata record (correlating to Archivematica AIP) that can be ingested to Archivematica as AIC record (= basically an AIP without the content) appended to a preserved AIP
18. PlumAnalytics. Does PlumAnalytics currently have, or is Plum planning, an integration with the DAMS?
19. VIVO. Does VIVO currently have, or is VIVO planning, an integration with the DAMS?
20. VIVO. Is there bidirectional communication through the triplestore between the DAMS and VIVO? [this does not exist?]
21. Handles/DOIs. What methods are available to create, delete, and edit handles or DOIs in the DAMS?
22. Handles/DOIs. What server configuration and firewall considerations are necessary for this integration?
23. Handles/DOIs. For DOIs, is there a feature that can express the DOI as an URI or is a separate HTTP proxy server needed?

Appendix 5: Simplified Testing Rubric

Tests were graded on a Yes/No/Partial scale.

- Robust API
- Broad adoption with community support
- Ability to function as a modular solution in a DAME setting
- Ability to serve as a standalone product
- Support for diverse file formats (including A/V)
- Support for complex objects
- Metadata versioning
- Provenance and authority checks
- Auditing capabilities
- Integration with Archivematica
- Integration with Shibboleth
- Integration with handle servers or DOIs
- Support for hierarchical and structured metadata for items, collections, communities
- Support for hierarchical and structured metadata for agents
- Support for configurable metadata
- Support for linked data
- Ability to authenticate and set granular permissions and restrictions
- Ability to upload, edit, publish, delete, and change permissions individually and in batch
- Ability to attach licenses to items, collections, or sets
- Ability to define complex relationships between items
- Discovery and indexing of items
- Ability to generate exportable reports
- Ability to track contributions
- Ability to track access/usage