

Title

Analysis of challenges and opportunities for migrating ScholarsArchive@OSU to a new technical platform: requirements analysis, environmental scan, and recommended next steps

Authors

Steve Van Tuyl, Data and Digital Repository Librarian, Oregon State University Libraries & Press

Hui Zhang, Digital Applications Librarian, Oregon State University Libraries & Press

Michael Boock, Head, Center for Digital Scholarship and Services, Oregon State University Libraries & Press

Executive Summary

ScholarsArchive@OSU (SA@OSU) has stood as Oregon State University's institutional repository for nearly a decade, and has seen a great deal of success, by many metrics, as such. Over that time period, the mission, content types, and stakeholders of the repository have changed, as has the ecosystem of available and emerging repository platforms. These changes have sparked interest in OSU Libraries & Press in assessing the technical infrastructures for SA@OSU to determine whether migration to a new system could benefit all of our stakeholders. This document presents the results of an investigation conducted by Center for Digital Scholarship and Services faculty and staff between September 2014 and January 2015 into the requirements of our stakeholders and the suite of practically implementable repository platforms for SA@OSU. We provide an assessment of a number of platforms in the context of our requirements analysis and make recommendations for next steps.

SA@OSU is currently hosted on the DSpace repository platform, as it has been since the repository's inception. The use of this platform has helped SA@OSU become the successful repository it is today, but there are elements of this technical environment that are restrictive or that prevent us from fully engaging with digital objects the way we would like to. One key example of these limitations is the strict organizational structure placed on a repository by DSpace, with a hierarchical and linear form of information organization (communities and collections) that prevents some modern types of interaction with the data in the repository. These restrictions have proven especially difficult in the face of continued proliferation of digital object types we collect, or could collect, in SA@OSU.

Our assessment of SA@OSU was conducted in three phases: determination of current functionalities of SA@OSU, a requirements analysis conducted with SA@OSU

stakeholders, and assessments, informed by the requirements analysis, of potential repository platforms. We make a recommendation for a new repository system, and offer recommendations related to repository function, administration, and expansion.

Our requirements analysis, a key component to our investigation, was conducted with OSU Libraries & Press administrators, faculty, and staff who work with or manage SA@OSU activities. We also included a number of external stakeholders who represent contributors and users of SA@OSU in order to understand, and incorporate, the needs of our end-users. Most, if not all, of the participants in our requirements analysis have a cohesive vision for the overall purpose of SA@OSU. While there is some diversity of opinion on the technical details of how this vision could be implemented, there is no significant disagreement that there are some major changes or features that would benefit the stakeholder community. Primary among these improvements are the need for a higher quality user experience, more flexible and reliable statistical reporting, and a technical infrastructure that allows SA@OSU to represent the complex structure and interrelatedness of digital objects. These, among other less technically significant functional needs point to the need to rethink the technical infrastructure of the repository in order to best meet the needs of our stakeholders and represent the digital objects in SA@OSU as authentically and usefully as possible.

To understand what platforms we might take advantage of to help implement a new version of SA@OSU, we assessed a set of mature repository platforms including DSpace, Hydra/Sufia, and Digital Commons. This environmental scan suggests that Digital Commons does not have sufficient functionality to meet our needs. Considering the other two platforms, we recognize that both DSpace and Hydra/Sufia are viable options for SA@OSU, but that many of the current concerns we have about our repository are due to restrictions imposed by the DSpace environment. While the adoption of Hydra/Sufia would mean a significant commitment of resources at OSU and from the wider Hydra community, we believe that a successful move to Hydra/Sufia would result in a repository platform that best meets our current ecosystem of digital object types, services needs, and potential future functionalities the SA@OSU may be called on to provide.

Implementation of Hydra/Sufia will not be a trivial task, and the first step in this process should be to engage with the Hydra development community to understand how our repository needs can be prioritized. Early conversations with influential members of the community suggest that the move of a large repository like SA@OSU to Hydra could influence the up-prioritization of OSU's project in the overall Hydra development stream. We recommend implementation of technical planning, metadata enhancement, and

preparation of our current store of digital objects as soon as possible. This should be followed by development of core functionality for the repository and migration of a subset of object types to the new repository environment. Following stress testing of the new system - we should complete development around the workflows required by our staff and faculty for deposit and management of SA@OSU content, along with a final migration to a new system (by 2016). Technical discussions and assessment of development needs will obviously affect the timeline of this project, but our goal is to move from our current platform to a new platform post-haste. Migration will, clearly, not be considered complete until we have a fully functional repository system in place.

Introduction

ScholarsArchive@OSU (SA@OSU) is Oregon State University's digital solution for capturing, preserving, and making available the scholarly content of our academic community. Across nearly a decade of use, SA@OSU, which uses the DSpace digital repository platform, has accumulated more than 53,000 individual items including faculty research papers, electronic theses and dissertations, and historical university Extension Office publications. The repository is regularly ranked among the top ten single institution repositories in the country based on the indexing of items in google scholar, number of items and related criteria. While the repository has long included a wide variety of formats, object types and research outputs, recent efforts to include research datasets has prompted CDSS to consider whether DSpace remains the best repository option for OSU, to reflect on the effectiveness of SA@OSU as a single repository, and to consider whether the proliferation of digital object types points to the implementation of a new platform for repository services.

This report presents the results of a requirements analysis and environmental scan conducted from September 2014 through January 2015 by CDSS. Our goal in this process was to assess the needs for our repository spaces and the types of content the repository can be expected to hold and serve now and into the future. To this end, we spoke with as many stakeholder groups as possible including library administrators, internal library users (primarily CDSS faculty and staff), and OSU community stakeholders. We also explored the current space of repository platforms focusing on emerging trends and needs in the academic repository environment.

Existing Feature Set of SA@OSU

Given the maturity and size of SA@OSU, it is important to understand the baseline that any migration or major change to the repository system will need to achieve. The

workflows and functionalities included in our current platform (DSpace 3.X) are necessary to retain with any migration; they are critical to the functioning of the repository and the staff that supports it. This current functionality should serve as a benchmark that must be matched by any future system. This list of functionalities described below is meant to be broadly written. Detailed elucidation of these functionalities (e.g. with user stories and use cases) may be conducted in the future as part of the decision-making and planning process.

Current Functions of DSpace

- Capture/Ingest - flexibility in modes of data capture including user submission and customizable workflows for submission (e.g. based on type of data being captured)
- Index - metadata are indexed in a manner that allows for retrieval and reporting
- Retrieve - interactive elements allow for discovery and retrieval of items in the repository based using multiple modes of discovery (e.g. search, browse, faceting)
- Preserve - basic preservation functionality including checksums
- File Agnostic - ingest, preservation, index, and retrieval can take place, at least at a minimum level, for any file format
- Access Controls - both internal (administrative) and external (e.g. embargoes)
- Reporting - reporting of usage information to multiple classes of stakeholders

Requirements Analysis

Beyond identifying current functionalities in our repository system, we sought to discuss the repository system with a number of stakeholders both inside and outside of the library to better understand conceptually and practically how SA@OSU is and isn't meeting the expectations of the user community. To this end, we identified three major stakeholder communities: administrators, internal library users, and external users. We created questionnaires for each of these stakeholder groups with questions meant to elicit discussion about what and who SA@OSU is for and how well it provides functionality to those communities.

Administrator Elicitations

For administrator elicitations, we spoke with the three library administrators who are directly involved with providing vision and guidance for SA@OSU and for the technical aspects of our scholarly communications landscape (Shan Sutton, Michael Boock, and Evviva Weinraub). There is a great deal of convergence from these administrators about the overall, conceptual purpose of SA@OSU (who it is for, what it should do, why it is important), though there is a bit of divergence around the types of content

SA@OSU should hold. A strict definition of “scholarly” is particularly difficult to pin-down, though administrators agree that scholarly content is the primary content type for SA@OSU and that the repository should maintain high standards for content quality. That said, the collection development policy (from 2008) and the content currently in SA@OSU seem to cast what could be interpreted as a slightly broader net for content allowable for the repository (e.g. including teaching materials, speeches, historical documents, etc.).

When discussing innovation in relation to the repository, administrators all suggested that it is important for SA@OSU to be as technologically innovative as possible without impeding functionality and sustainability. The same applies to service functionality - administrators are adamant that SA@OSU (and CDSS by proxy) should be very innovative in the types of service we offer around the repository as long as CDSS is able to provide personnel and time for support of these services. There is, in reality, a good deal of overlap between services offered and technical capabilities of the repository and clearly identifying and prioritizing technical needs for repository services will be important in any new iteration of SA@OSU. An example of this is the implementation of Altmetrics in SA@OSU - the technical implementation is crucial to providing this service, but identifying how this service meets faculty needs and promoting the service are in some ways equally important to the implementation of the technology.

There is also administrative convergence on technical details and requirements for SA@OSU. Administrators have relatively little disagreement about the types of technologies and tools that are important to the functioning of the repository. There is strong agreement around such technical elements as incorporation of current and emerging identifier schema (e.g. ORCID, FundRef), functionality for linking among objects in the repository, and the need for SA@OSU to be able to preserve and make available as many file types as possible.

Divergence on technical issues among administrators is often due to a sense that we need to fully understand the implications and benefits of committing to a technical element of the repository. For example, commitment to complying with various audit frameworks (e.g. Data Seal of Approval) is a priority for some, but others are supportive with the caveat that we understand what we get by working towards compliance. Other examples of divergence are around whether the codebase for the repository should be open source and whether the repository should have an available API.

Internal Library Elicitations

The internal elicitations were conducted with members of the Center for Digital Scholarship and Service who regularly interact with the repository either directly or through management of projects and personnel using the repository. These elicitations included Sue Kunda, Amanda Whitmire, Andrea Wirth, Brian Davis, Maura Valentino, and members of the Scholars Archive Users Group (Erin Clark, Deanne Bruner, Patricia Black, and Laura Wilson).

Similar to the responses from administration, there is widespread agreement by internal library users as to the purpose of SA@OSU, what it does well in its current instance, and how it might be improved in the future. Users who directly interact with SA@OSU on a daily or weekly basis tend to have more specific comments about what the repository could do better in future iterations. These improvements can be split into two categories: issues with the functionality of our current platform (DSpace 3.X) and issues with auxiliary systems we use for the repository (e.g. Solr, web interface). With respect to the functionality of our version of DSpace, the primary points of concern are around the lack of hierarchical metadata, the structure that DSpace forces on objects in the repository (Communities and Collections), the need for more flexibility in workflows for ingesting and editing content, and the system's inability to natively provide links among objects in the repository. Issues with auxiliary systems include the need for a more "updated" or "professional" look and feel, the need for highly flexible statistics and reporting mechanisms, and the need for external controls on content such as name authorities and cleaner metadata.

When asked about specific technical needs for a future version of SA@OSU, responses from internal users converged on a few key areas that were also prioritized by administrators. First, the need for more flexibility and personalization in reporting on SA@OSU content (usage statistics, page views, uploads, etc.), both for internal and external users, emerged as a primary need in almost every elicitation. Second, the look and feel of the repository was criticized as not seeming professional or up to date, though interviewees rarely were able to identify, specifically, what issues were problematic. Last, the ability to create active links between objects in the repository and the related need to represent complex relationships among objects in the repository is seen as crucial to accurately representing the content within.

It is important to note that many of the functionality issues expressed by the Scholars Archive Users Group are relatively minor improvements (e.g. allowing diacritics) but the sheer number of those improvements add up to a fairly sizeable set of problems for the repository. One should note, too, that this group is far and away the primary user-base for the repository software given that they interact with the system on a full-time basis. It

is crucial that any new system address the needs and requirements of this group as completely as possible in order for the system to be considered an adequate replacement and improvement for SA@OSU.

Finally, a few of the internal elicitations generated detailed lists of functionality that could greatly improve the repository for internal and external users (see Appendix A). These recommendations should be revisited as we move further into the planning and development stages of this project.

External Elicitations

The needs and concerns expressed by external users were widely varied and often very specific to the way the user group interacts with the repository. One common thread throughout the external elicitations and other informal discussions with external stakeholders is that it is crucial for SA@OSU to provide consistent, reliable, and clear statistical reporting. Almost all external users expressed concern about the lack of flexibility in reporting out of SA@OSU and frustration with practices such as retroactively changing download counts (through the process of cleaning data of bot/crawler traffic).

Environment Scan

We have the following criteria when looking for platforms of future SA@OSU: feature set, support and sustainability, openness, and popularity. There are, essentially, three major repository platforms that we might consider for a repository of the size and complexity of SA@OSU: DSpace, Hydra (Sufia), and Digital Commons. This is not to say that these are the only players; rather, they represent the most likely platforms of three different types: open source “turn-key” solutions (DSpace), open and customizable frameworks for a variety of repository services (Hydra), and hosted solutions (Digital Commons).

There are other platforms that fit into these categories including such solutions as EPrints, an open source solution similar to DSpace, and Islandora, a software stack similar (in some very loose ways) to Hydra. In order to narrow the field of options, we’ve selected the three mentioned above as representative of the types of solutions we could choose. The selection of DSpace is obvious, as it is the current platform for SA@OSU and is a relatively dominant player in the market. Similarly, the selection of Digital Commons is driven by their relative dominance in the hosted solution space. The selection of Hydra is driven by the growth of the development community over the past

few years along with personal experience and technical expertise with the system at OSU through the Oregon Digital project.

DSpace

DSpace is an open source “turn key” institutional repository platform that has risen in popularity over the last decade. The DSpace community boasts the flexibility of the platform to meet a variety of needs and the platform has a proven long-term track record. There are currently over 1000 institutions running DSpace and there is an active community of developers making updates to the platform regularly. DSpace is the current and has been the only platform for SA@OSU.

Available Features

- mediated deposit
- batch metadata editing
- versioning
- usage statistics
- optimized for Google Scholars
- OAI-PMH
- SWORD

Development Needs

- support complex object
- link between object
- batch upload and editing
- intermediated deposits
- flexible collection structure
- flexible metadata schema including hierarchical support
- integration with cloud storage

Long-term Sustainability

DSpace has been installed in more than 1,000 institutions worldwide and its development is also subsidized by Duraspace. With a solid developer community and large user base, DSpace will continue to be a top choice of digital repository in the future.

Recommendation for DSpace

DSpace is a robust and well supported repository system - as evidenced, in part, by the system’s ability to handle the inception, growth, and diversification of SA@OSU over the past decade. One of the major benefits of DSpace over other platforms is our familiarity

with the system and the (potential) relative ease of transitioning to a new version of DSpace. That said, we are keenly aware of the limitations of DSpace especially as we increasingly consider new content types, stakeholders, and modes of interaction with the repository. DSpace is certainly still a viable solution for a repository platform for SA@OSU, but we feel that we cannot remain with DSpace without putting a great deal of effort into revising the structure of the repository and considerable development work to meet stakeholder needs.

Hydra/Sufia

Hydra is a repository framework that is developed and managed by a multi-institutional community of developers and users. Hydra is currently supported by developers at about 25 institutions with an additional 20-30 institutions running Hydra for digital asset management. The technology supporting Hydra includes Fedora, Solr, and Blacklight along with a great deal of community developed work to make these pieces function as a single unit.

Sufia is developed as a component (Rails engine) of a Hydra framework that adds features of a self-deposited IR. It was created at Pennsylvania State University libraries and has attracted interest from the Hydra community for further development. Based on the recent update from Hydra Connect conference in Fall 2014, there is a plan to merge Sufia with several similar projects and extend its feature sets starting the second quarter in 2015. So far, institutions such as Stanford University, Pennsylvania State University, and Cornell University have already expressed interest in the project. We have also heard strong verbal commitment from members of the Hydra community for a large amount of development aimed at broadening the applicability of the Sufia framework.

Available Features

- batch editing and upload
- full text search
- fixity checking
- version control
- usage statistics by Google Analytics
- integration with cloud storage
- optimized for Google Scholars
- proxy deposit and transfers of ownership

Development Needs

- mediated deposit

- complex objects (e.g., child and parent objects)
- link between objects
- flexible collection structure
- flexible metadata schema
- administrative interface
- support various file format and size
- OAI-PMH
- embargo
- SWORD

Long-term Sustainability

Hydra projects, such as Sufia, are by nature collaborative efforts from various partners. Hydra is not supported by any grant or membership fees, it depends on contributions from communities (Hydra partners and developers) in codes, ideas, and documentations. While relatively new to libraries, the community-based development model is a long-lived and successful model in many other industries. Any concerns about the longevity of the Hydra community should be allayed by the fact that the community has plans to develop long-term strategy and sustainability plans this year, and that the community's foundational technologies (e.g. Fedora) have viable sustainability plans in place (e.g. DuraSpace).

The flexibility and openness of the community centric approach to development are two of its major advantages. All the development codes are available as open source and each institution is free for customizing or extending the code bases for its own requirements and has the obligation to share its efforts with other partners.

Furthermore, Hydra community nowadays includes some major research institutions (e.g., Stanford, Penn State, Indiana, and University of Virginia) as active partners and it is still growing. Hydra has a proven record of success since 2008 and will attract more contributors in the foreseeable future.

Recommendation for Hydra/Sufia

Hydra/Sufia could provide a highly customized (and customizable) repository framework for SA@OSU. Moving to an entirely new platform could allow us to radically rethink our workflows and representation of digital content in SA@OSU, and the relative newness of the platform presents many opportunities for building the type of repository our stakeholders have expressed interest in. That said, Hydra/Sufia isn't necessarily the number one solution for SA@OSU and adoption of this platform poses a number of difficulties, not the least of which is development work needed to bring the system up to our expected standards.

Digital Commons

Digital Commons is a hosted repository solution licensed by bepress with modules for institutional repository functionality, publishing, and system for creating faculty profiles. Digital Commons is used by hundreds of institutions for a variety of repository needs including faculty and student publications, theses and dissertations, and multimedia file hosting.

Development Needs

The extent to which Digital Commons offers the opportunity to customize individual repository instances is quite unclear. Most Digital Commons instances have a very similar look and feel which suggests a fairly static set of options for customization. The concern with a hosted, static user interface for the repository is that any modifications we need to make are either intermediated and vetted by the hosting organization (bepress in this case) or are workarounds for an inflexible hosted solution. In addition, working in a hosted environment means that managers of SA@OSU lose the ability to promise consistent uptimes, updates to codebase, and other assurances that stakeholders have come to expect. That said, we have not sought direct input or comment from bepress about the modifiability of their system to meet SA@OSU needs due to other concerns about the appropriateness of the platform for our needs.

Long-term Sustainability

bepress has been licensing Digital Commons for about a decade and the system is used by hundreds of institutions worldwide. We see no indications that bepress or its support for Digital Commons is in danger of folding or otherwise losing support. Our primary sustainability concern with Digital Commons is that it is not clear how committed bepress is to providing support for “non-standard” digital objects (e.g. datasets). Personal communications with outreach and support personnel at bepress about their plans for working with datasets failed to provide confidence in the degree to which bepress is taking a forward thinking role in the area of storage of “non-standard” objects.

Recommendation for Digital Commons

There are three primary difficulties we see with using the hosted Digital Commons framework: 1. concerns about support for non-standard digital objects, 2. a lack of support for complexity in representing digital objects, and 3. a the lack of support for hierarchical metadata. Our concerns around support for non-standard digital objects have been expressed in the Long-term Sustainability section above. Support for complex relationships and hierarchical metadata are major concerns for SA@OSU and

have come up repeatedly in our requirements analysis with stakeholders. The framework offered by Digital Commons does nothing to resolve this problem and, in many ways, replicates the environment that we are currently working in (flat metadata, linking to related files without representing the nature of the relationship between files, etc.). The lack of support for hierarchical metadata, is directly related to the first issue and results in an inability to represent relationships among digital objects and the relationships between and among metadata elements for the object. A key example of a problem the flat nature of the Digital Commons (and our current) metadata framework causes is our inability to link authors to related identifying information in our metadata records (e.g. author identifiers, home departments, institutions, etc.).

Our recommendation is that Digital Commons is highly unlikely to meet the needs of SA@OSU, especially with respect to new, high priority requirements that have emerged from our requirements analysis.

Final Platform Recommendation

Regardless of our choice of platform, it is important for us to re-think the purpose of the new repository and re-design its features to meet the user demands. While the current repository platform is functional, we recommend a complete migration to a new system (DSpace or Hydra) including a complete overhaul of our metadata, workflows, and information architecture.

Many of the decisions about the structure and function of SA@OSU have been made as a reaction to restrictions imposed by DSpace. Maybe more important, the workflows, metadata practices, and policies we use for SA@OSU are the result of over a decade of initial planning for, development in, and upgrade to the DSpace infrastructure (by OSU and by the wider DSpace community). Our current systems are also the result of over a decade of reactions to unique cases brought before the repository including new programs and initiatives at the university, calls for the expansion or modification of services (e.g. modifying statistical reporting, initiating work with Extension and Experiment Station Communications), and calls for differential treatment of materials (e.g. indefinite embargoes on theses from some departments). These are problems experienced by any large-scale system operating in and providing service to a dynamic environment like a research university - periodic, major review of these systems is not uncommon and might be considered best-practice.

The communities and collections structure of the repository, for example, poses challenges for offering complexity and nuance in interactions with the system by fixing content into that hierarchical structure. While there are resolutions to this hierarchy

within DSpace (e.g. mapping items into multiple collections), these solutions rely entirely on the relatively inflexible DSpace architecture. An alternative approach to representing the structure of our communities and collections, which is not available to us in our current platform, is to rely on object metadata to dictate relationships among objects and entities (e.g. relationship between a research paper, an author, and an academic department), which would offer more flexibility in information retrieval, display, and structural representation.

A second major example (though related to the above) of how an overhaul of the system could benefit our stakeholders has to do with taking a new approach to metadata for our digital objects. The metadata schema used for DSpace - largely Dublin Core - is flat in nature which makes it difficult to represent hierarchies in relationships among objects. Currently, relationships between and among objects can be represented in a non-active way, with notes in metadata representing relationships, but without active links. Our current form of metadata also lacks functionality for hierarchical representation of metadata which prevents us from representing rich relationships within the metadata such as the relationship between an author and that authors ORCID or the relationship between the item and the funding sources for the research.

Stakeholder requirements don't strongly specify one solution over the other. They do, however, point towards the need for a level of customization in the repository that we have not been able to achieve in the past few years. This indicates to us that, regardless of platform, we need to commit a great deal of thought and resource to reconceptualizing the repository platform to meet the varied needs of our stakeholders and to show value to our community. This includes creating more useful and usable interfaces generally (our user experience), offering flexibility in reporting far beyond what we provide now, and considering the variety of ways we can and should represent the digital objects in the repository (e.g. compound digital objects, linking among related objects).

We must also consider the infrastructure needed to maintain critical workflows used by many of our users both internally and externally including customizable forms, access controls, and permissions. A migration cannot be considered complete if these functions are not incorporated and replicated at least at the level of functionality currently experienced by our users, but we also strongly suggest that any migration include improvements to these functions - simply replicating the current infrastructure would not constitute a worthwhile migration.

With these points in mind, we believe that the most productive course of action would be to move towards a Hydra-based repository system. Given the relative newness of the Hydra community and the extent to which repository functionality in Hydra-based systems (e.g. Sufia) are underdeveloped, there are significant challenges and opportunities to making this choice. The obvious primary challenge is that the lack of a developed infrastructure around Hydra for institutional or other repository services means that a great deal of work will need to be done by the OSU community and the broader Hydra community to create the level of functionality we require. Early conversations with leaders in the Hydra community (specifically Michael Giarlo at Pennsylvania State University) have resulted in tentative commitment to begin development for at least some of these services in the coming year - especially if a large repository like SA@OSU is involved. We may, however, need to more formally secure commitment of prioritization and development time on the part of the community (as much as is possible).

Assuming commitment of time and resources is available to rebuild SA@OSU in a Hydra environment, we can identify a number of opportunities that emerge from working in such an underdeveloped space. First, this offers an opportunity to rethink current workflows and needs and determine what workflows are critical and what workflows are simply artifacts of the repository platform we currently use. Second, services and functions, such as statistics and reporting, that do not currently meet our needs (or do not meet our needs well) can be completely rebuilt with a focus on the needs of our users. Last, committing to work in the Hydra environment will leverage current development expertise and experience in the library and contribute towards a unified development environment.

Other Recommendations and Considerations

The discussion about selecting a new repository platform necessarily raises questions that are important but not dependant on the final outcome of the platform decision making process. We would like to take the opportunity to note a few of these questions and issues that have arisen during the process of speaking with stakeholders and thinking about a new vision for the repository because we believe that addressing these issues will affect our vision for SA@OSU.

First, from discussions with stakeholders at all levels, it is obvious that there is some confusion about the scope of materials that should be collected in SA@OSU. All stakeholders agree that “scholarly” materials are the primary focus of the repository, but some stakeholders, along with some documentation for the repository (e.g. the 2008

collection development policy) indicate a number of other content types that could be included (such as teaching materials and undergraduate scholarship). We also see a not insignificant number of collections and item types in SA@OSU that do not fit the model of “scholarship” but clearly are in need of repository services (e.g. data from OSU Institutional Research). It will be helpful to come to broad and well documented agreement as to the types of content that do and do not belong in SA@OSU.

Related to this first point is the fact that there are currently types of content that we do not cater to in SA@OSU either because they do not fit our current collection scope or because there has been little call for repository services around these content types. If we determine that some content types are not appropriate for SA@OSU, and those types of content call for repository services, we need to also engage in discussion about whether the library can or should provide those services.

Two external stakeholders very clearly indicated that they thought OSU Libraries & Press did a poor job of showing the value of SA@OSU to faculty, especially for non-STEM research domains. We feel that it is important from a service perspective to try to better understand why researchers in non-STEM domains down-value the services provided around SA@OSU and whether we can or should try to expand or market services to these fields specifically.

Last, there are a number of administrative documents and policies that are missing or need to be updated for SA@OSU. These include documentation about preservation planning, succession planning, and our retention policies. These types of administrative documents serve many purposes including allowing SA@OSU to be certified as a trusted repository, clarifying our preservation commitment to stakeholders, and offering a degree of clarity and direction for the long-term management of the repository.

Appendix A.

New Functionalities or Improvements to Functionality as Presented by Participants in Requirements Elicitations

Ingest

- Customizable workflows
- Increased level of automation for deposit (e.g. reading metadata from Web of Science feeds, deposit forms, etc.)
- Batch ingest of content
- Pull data from CrossRef

Information Retrieval

- Authority files
- Support for diacritics and other character sets as they emerge
- Search and browse functionality that is clear and helpful to users
- Integration with external resources such as SHERPA/ROMEO, ORCID, etc

Statistics and Reporting

- Author Profiles
- Statistics and reporting that are reliable and consistent
- Flexible reporting functionality for internal and external users

Metadata

- Hierarchical metadata
- Metadata used as the primary mode of sorting and grouping information
- Batch editing of metadata

Linking

- Active links between objects
- Support for compound digital objects
- Need to be able to link to content outside of the repository

Access Controls

- More granular bitstream level administration (e.g. bitstream level policies)

User Experience/Design

- More professional look and feel
- Visualization of information in the repository through thumbnails, streaming AV, previews

- Representation of the structure (e.g. file structure) of multi-part objects

Appendix B.

Tentative Milestones for Recommended Hydra Solution

Development and implementation of a new repository system is a large project. Early discussions between CDSS and ETS suggest the following phased approach to repository development and migration. The phases and milestones described below are abbreviated for the purposes of this document. Additional information is available for each of these phases and milestones, though, as with any large-scale project like this one, we expect that a degree of flexibility should be built in to both our proposed approach and our timeline. That said, this is our best estimate, at this early date, for the milestones of this project.

Phase I - To be completed by Summer 2015

Milestone I: Phased Development of Data Repository

Development of basic repository functionality (ingest, security, authorization, metadata cleaning/decision making):

- Stand up a repository for a small but working subset of SA@OSU content (primarily datasets)
- Leverage existing frameworks for Hydra-based repositories (e.g. Pennsylvania State University's ScholarSphere and the spinoff Sufia)
- Initiate wholesale metadata cleanup and preparation for migration to new repository system

Milestone II: Implementation of Multipart File Functionality

Implementation of functionality for linking among files in the repository and representing files or file groupings with multiple parts

- Display Hydra::Works application profile for multipart items (linking, file hierarchy, etc.)
- Implement improved statistics reporting beyond basic Sufia functionality
- Migrate subset of multipart and linked objects to the new repository system (primarily datasets and linked documents)

Phase II - To be completed by late winter 2016

Milestone III:

Implementation of mediation processes for repository submissions.

- Development of infrastructure allowing mediated deposit
- Implementation of additional reporting functionality including author profiles
- Improve functionality around batch ingest and batch metadata editing

- Planning for training of repository users (internal and external)
- Planning for marketing and outreach for new repository

Milestone IV:

Complete migration of content from DSpace to the new repository system.

- Full migration of content and shut down of DSpace
- Training of users on new repository system, workflows, and functionality
- Initiation of marketing and outreach for new repository