Against the Grain

Volume 29 | Issue 2 Article 69

April 2017

Biz of Acq-Implementing MD-SOAR, a Shared Consortial Repository

Michelle Flinchbaugh
University of Maryland Baltimore County, flinchba@umbc.edu

Recommended Citation
Flinchbaugh, Michelle (2017) "Biz of Acq-Implementing MD-SOAR, a Shared Consortial Repository," *Against the Grain*: Vol. 29: Iss. 2, Article 69.
DOI: https://doi.org/10.7771/2380-176X.7768

Biz of Acq — Implementing MD-SOAR, a Shared Consortial Repository

By Column Editor: **Michelle Flinchbaugh** (Acquisitions and Digital Scholarship Services Librarian, Albin O. Kuhn Library & Gallery, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250; Phone: 410-455-6754; Fax: 410-455-1598) <flinchba@umbc.edu>

Column Editor's Note: *Acquisitions units are taking on work to support digital collections and intuitional repositories, which I refer to together as "digital repositories." While acquisitions can support repositories by acquiring digital content, conducting quality review of digital content, moving digital content between systems, and inventorying, manipulating, and ingesting digital content into a repository, experience in working collaboratively in a consortial environment can also position acquisitions librarians to lead collaborative digital projects. The February 2016* ***"Biz of Acq"*** *column featured an article,* ***"MD-SOAR, Maryland's Shared Open Access Repository: It's been a Long, Long Haul"*** *on the work necessary to move an IR concept from an idea to a pilot project for a shared digital repository. The two-year pilot project for implementing MD-SOAR (****https://mdsoar.org/****) began on April 1, 2015, and this article covers the implementation process. — **MF***

The Maryland College Shared Digital Initiative (MDCSDI) moved from planning for a shared institutional repository to the implementation phase on February 1, 2015. The group agreed to implement the repository on the DSpace platform (*http://www.dspace.org/*), to be hosted by Digital Systems and Stewardship (DSS) at the **University of Maryland, College Park**, and obtained funding from the **University System of Maryland and Affiliated Institutions' (USMAI)** Council of Library Directors. They had also established which Maryland colleges would participate in the pilot: eight **USMAI** libraries plus **Goucher College**, **Maryland Institute College of Art**, and **Loyola Notre Dame Library**, which joined the **USMAI** consortium during the pilot. With the first implementation meeting, MDCSDI became known as the Governance Group.

While the two-year project wasn't divided into parts, there have been four distinct phases, which I'm naming and utilizing to organize this article: 1) Pre-implementation; 2) Implementation; 3) Post-implementation; 4) Evaluation and planning. For pre-implementation, the Governance Group's work fell within three major areas: infrastructure, implementation planning, and policies. During implementation, the Governance Group's work focused on customization and configuration decisions, loading, and support & training, while completing policies. Post implementation, after the libraries started to use the newly implemented repository, the group worked on enhancements (Creative Commons Licenses), reports and statistics, and usability. The final phase of the project, evaluation and planning for the future, is now in progress.

Pre-Implementation

Infrastructure — Infrastructure issues focused on how the group would work and communicate, and how the group and individual members would communicate with DSS. We had decided that the Governance Group would function democratically with each library getting one vote in decisions impacting the platform. We also had one contact from most libraries participating. We immediately asked each library to additionally name alternate contacts to ensure that all campuses were aware of key issues as implementation progressed. We determined that all meetings should be open, so that specialists not on the official contact list could attend either as substitutes or in addition to regular members to provide input into discussion and decisions. Email lists, which had been hosted by **UMBC**, were migrated to the host site at the **University of Maryland, College Park**. The group's Web page was migrated from **UMBC** to Basecamp, a Web-based project management and collaboration tool (*https://basecamp.com/*). Later, when libraries had trouble finding relevant policies in Basecamp, policy documents were moved to a MD-SOAR Web page on the public **USMAI** Website, along with a list of campus contacts. In addition to organizing the governance group, we also had to determine how the group would work with DSS. DSS named contacts who we would work with throughout the project. With feedback from the group, the **USMAI** Executive Director and DSS drafted a "Service Level Agreement" outlining the services that participating libraries would receive. The Service Level Agreement was between DSS and **USMAI** rather than between DSS and the individual libraries since **USMAI** provided 100% of the funds for the project.

It's important to note that while some elements of the infrastructure were set, there was a great deal of flexibility in how we went about making decisions. Workload stress was an issue that always had to be taken into account in figuring out how to get things done. Most issues were worked on by a small group, which would submit a plan or policy draft, for discussion, possible modification, and vote. Sometimes, during group discussion, a plan would emerge, and barring any objections, would be accepted. As metadata is complex issue, and the Governance Group had only two members with expertise, we delegated it to a standing sub-group with additional members with appropriate expertise, and gave that sub-group decision making authority. In the instance of record displays, there were very strong opinions on a very detailed level, so the sub-group working on the issue submitted two possible plans — the group voted on the plans, then each library proposed modifications and the group voted on each proposed modification. In the instance of usability, a usability study was delegated to a **USMAI** User Experience group. It's important to note that Governance Group members by-and-large were responsible for their library's implementation of the repository along with the duties of their regular full-time job, and depending on their current workload or projects in their library, were not always responsive or engaged in the decision-making process. Essentially, there was no right way, but rather, a variety of different methods needed.

Implementation Planning — The first implementation decision the group had to make was a consequence of implementing a single, centrally-hosted system for all of the libraries to use. There would be only one URL for the site, so libraries would not be able to use their own URLs for it. After some discussion, the group agreed to call the repository MD-SOAR (The Maryland Shared Open Access Repository), and to base the URL on that name. Further, the **USMAI** Executive Director agreed to hire a graphic designer to create an MD-SOAR logo to appear on the site. Each library would have a community within the repository, which could contain limitless collections and sub-communities. After some discussion and research on the part of DSS, the group agreed that each library would also provide a university logo to appear on all the pages within their community for continuity in university branding.

In advance of the first implementation meeting, on the request of a participating library, the **USMAI** Executive Director, the Director of Consortial Library Application Support in DSS, and the Governance Group Chair agreed that the first thing DSS would do was set up a sandbox DSpace site to allow participating libraries to become familiar with the software. Libraries were given access to the sandbox site at the first implementation meeting. In addition to the sandbox site, a staging version of the software would be set up, in addition to the live version, for testing both loads and interface changes before making them in the live version of DSpace. The sandbox site was eventually taken down.

Policies — During the first implementation meeting, the Governance Group reviewed repository policies from other schools, then determined what polices would need to be developed for MD-SOAR: a file-size policy, a content and file format policy, a metadata policy, and a take-down policy. A file size policy would address limitations on storage; with subsequent discussion, the group determined to wait for problems to occur before addressing this concern. Thus far, none have occurred, so a file size policy was never drafted or adopted. The group immediately began work on a content policy and metadata policy, assigning two group members to work on both of those tasks. The group also agreed

continued on page 60

to work on a license agreement at this time, and one person agreed to adapt the existing **University of Maryland, College Park** repository license for the group. The license agreement was adapted with few issues and little discussion, but with the understanding that each participating library would consult their campus legal counsel, making the identified agreement a template to be modified by each campus as mandated by their individual counsel. The take-down policy was put off until after implementation, since it was not needed in advance of implementation.

The content and format guideline was drafted, and readily adopted, after expanding scope in several areas to allow all libraries to use the platform as they wanted. In the first draft, the policy states that all items in MD-SOAR must be open-access, but some libraries wanted to limit access on certain items so this was modified to allow restrictions based on the needs of participating libraries. The first draft limited the scope to works by current faculty, staff, students, or academic or administrative units, but was later expanded to include current and former people of those categories, so that emeritus faculty could participate. This would also free libraries from having to remove works after an author left the university. The initial draft stated that items should be scholarly or academic in nature; this was modified to include part of or related to existing library collections, which was important to libraries planning to use MD-SOAR as a platform for digital special collections. The final policy is available for viewing here: *http://usmai.org/sites/public/files/ContentandFormatGuidelines.pdf*.

The take-down policy was also adopted only after expanding its scope. The initial draft included the most common instances, such as copyright violation. Research subjects with personally identifiable information revealed were added to the policy, as were agencies with authority over the work in whole or in part. The host university or department was given the right to remove student work that doesn't meet their quality standards. Beyond a policy for what would be removed, the group also had to develop a process for handling take-down requests. This required both standards as well as flexibility to reach an agreement. We needed a policy that would allow for responsiveness when campuses are understaffed and unresponsive to shield the group as well as the host from lawsuits; however we also needed to allow each campus discretion over its own works. There were a variety of opinions as to what to do once the determination was made that there was a problem with an item. The group decided that all take-down requests would go to DSS, which will forward the request to the campus involved. The campus is then given seven days to respond, and if no response occurs, DSS will remove all access to the item until the issue is resolved by the host campus. While called a take-down policy, the group determined that campuses at their own discretion could determine to remove a work entirely, move it to a dark archive by putting view limits on it, or modifying the work by removing a problematic portion (with a note in the metadata indicating that the change had been made). The final Take-Down Policy is available here: *http://usmai.org/sites/public/files/TakeDownPolicy.pdf*.

In repositories, it's common to organize materials roughly by organizational structure, so that each academic department has its own collection. With many universities sharing the same repository, we quickly realized that we were likely to have multiple collections all with the same name that are indistinguishable from one another. For example we might have eight history department collections. In DSpace, the collections appear in searches, at the top of the results, so having multiple indistinguishable collections all with the same name didn't make sense. Because of this, the group determined to include a campus prefix in all collection and community names. This, however, is a soft policy, in that if a collection has a name that is clearly and truly unique, the prefix can be omitted. For example, a collection might be **UMBC** History Collection, but the **UMBC** wouldn't have to be included in an **Albin O. Kuhn Library & Gallery** Collection, but this is ultimately up to the campus.

The metadata policy was by far the most complex and time-consuming. Also, the Governance Group only had two members with expertise in this area. After an attempt at a simple policy failed to work with DSpace because of misconceptions about system functionality, a metadata subgroup was formed with two members from the Governance Group, and two metadata librarians not on the Governance Group. Many hours of discussion went into developing this policy, available here: *http://usmai.org/sites/public/files/MD-SOAR_MetadataPolicies_rev_08_20_2015.pdf*.

Implementation

The live MD-SOAR server was set up by DSS. Important benchmark dates were the system go-live date, and when participating libraries received the go-ahead to begin submitting materials, several months later. During the implementation phase, a Staging server was set up that would serve as a permanent testing site to preview software upgrades, configuration, and loads. When the system went live, server work on it began happening on a release schedule, so that changes to live MD-SOAR only happened periodically, and only after having first been previewed on the Staging server.

Customization/Configuration — The Executive Director of the **USMAI**, the project funder, agreed to hire a graphic designer to design a logo. After discussion, the group agreed to use a mortarboard and the state flag in the logo. The group received back four possible logos from the graphic designer, discussed, asked for some changes, and voted on them. The graphic designer finalized the logo, and it was added to the system. Each participating library also provided a campus logo, and these were all added to each libraries' individual community in DSpace. Individual library contact information was also added to the footer of each campus's community.

Upon finalizing the metadata policy, the Metadata Group wanted to customize the DSpace metadata drop-down menu to match the policy, hiding elements that were not adopted in the identified schema. However, DSS was concerned that the software use some of those elements. Additionally, they were concerned that we would simply want removed elements added back in later, especially if we added new libraries that needed those elements. However when the Metadata Group made decisions to customize the indexing, the "do not use" elements were not included in the indexes; so while the software continues to allow their use, they won't be indexed if anyone does use them, so adding a new library that will use them requires expanding the indexing to include them. These were the metadata and indexing customizations that could be agreed upon.

The Metadata Group also customized the submission form. At some libraries, there was a great deal of debate and a desire to have campus-specific customized submission forms, up until DSS stated that only one submission form is covered by the current contract, and that adding more would require paying a fee for extra customization. Facing additional cost, interest evaporated. The one submission form broadly covers most materials but provides no opportunity to include campus, format, or subject specific information. Campus information could, however, be added via templates that the libraries can create to add metadata elements to all of their records as they come in. The group decided not to allow embargoing via the submission form in the spirit of open access, and this issue has caused problems for libraries which must first enter an item via the submission form, making the item available to the public, and only add the embargo after that. With one form, in serving the needs of the many, some simply haven't had their needs adequately met, so this is an issue that will perhaps be revisited in the future.

The Governance Group formed a small group to work on the customization of short item displays. It turned out that participating libraries had very strong opposing opinions on display, with some wanting the short item display to be very short with few metadata elements included, and others wanting it to be very long with nearly every metadata element included. The small group ended up putting forth both a long and short version to vote on. The short version won, but each library was given an opportunity to propose additional elements to add to it. Each proposed addition was voted on, resulting in a comprise medium length short item display.

Loading — All libraries were given the opportunity to load materials into MD-SOAR. At first this was thought to be a one-time start-up activity, but with discussion, it became clear that some libraries would need to load materials, such as electronic theses and dissertations, on an ongoing basis. DSS provided instructions on preparing loads. Most libraries provided files as well as text file containing the metadata formatted appropriately for DSpace. However, with only this

continued on page 61

information, items could only be loaded into one collection. Libraries were also given the opportunity to run a program, which reformats files for load, and provide a collections file to map items into more than one collection. DSS set up Box accounts for each library to transmit files to be loaded to them. Loads initially go into a staging server, which gives the library an opportunity to check and make corrections before loading to the live repository.

Support & Training — The Governance Group was given a quick tour of the sandbox server as soon as it was set up. During every meeting for approximately the first six months, time was dedicated to question and answer. Many questions focused on how to do certain activities in DSpace, and loading. Information was posted in Basecamp, and additionally many questions were asked and answered there. In the summer of 2016, after the live server was available, the group hosted a half-day training session for any staff in participating.

Post-Implementation

Enhancements — Many enhancements were mentioned at one time or another by various group members during the implementation cycle. These possible enhancements included integration with campuses' single sign on, an inline video viewer, support for multimedia, various types of campus customizations, and the implementation of Vireo to support ETD submission. The pilot contract didn't provide funding for such enhancements, and no one wanted to ask for additional money until the pilot was successfully completed, so none were pursued. However, if the pilot proved successful, enhancements with wide support might be funded in a new funding cycle.

The one enhancement that could be provided immediately was the integration of Creative Commons licenses in the submission process as DSpace already had this built in, and the feature simply needed to be activated. This turned out to be challenging when options had to be customized, and help information provided for system users. The process extended over several months as configuration was determined and additional use guidance added.

Reports & Statistics — On initial implementation, built-in DSpace statistics were available to administrators, but fell far short of a group wish-list of statistics. The systems statistics were made available to the public, and Google Analytics and Tag Manager set up to run on the site with each campus given access for their site. A **USMAI** training session on Google Analytics gave campuses an opportunity for hands-on learning to use Google Analytics.

The Governance Group also looked at statistics provided by a third party vendor for DSpace. Despite providing additional analytical information not captured by Google Analytics, this approach was not fully implemented and determined to be cost prohibitive. After some discussion, the group was unwilling to ask for financial support for this approach when several customizations might be a higher priority. This decision was shelved and will be revisited at a later date.

Usability — Various disagreements occurred over platform customizations and wordings. With no clear way of assessing, the group decided that a usability study of the site might provide greater insight on its design. They asked a standing **USMAI** User Experience group to evaluate the site. The User Experience Group agreed to do this, and the Governance Group provided scope information on what to include in the study. After a few months, a lengthy report was provided with problems encountered and suggested improvements. Most were acted upon, resulting in an overhaul of the site's main landing page, as well its menus, and some other miscellaneous tweaks to improve the site's usability.

Sharing Promotional Materials — All participating campuses, as well the **University of Maryland, College Park**, a non-participating partner and server host, agreed to share promotional materials they had developed. Several campuses loaded materials in Basecamp, resulting in a stock of materials that could be used as is or re-purposed by others.

Evaluation and Future Planning

At this time, the MD-SOAR Governance Group is in the process of evaluating the project and planning for future support and administration. Obtaining ongoing funding requires documenting the success of the project and developing a payment plan that participating library directors will agree to. Additionally, funding for enhancements requires building consensus around them, projecting their cost, and including that cost in the upcoming request for ongoing funding. The Governance Group has additionally compiled a list of achievements, and will provide statistical data to document success, such as the number of items uploaded, and the number of visits to the site. All participating libraries have been surveyed about their satisfaction with MD-SOAR, and future needs, including what customizations are considered critical and highly desirable. Participating libraries were additionally surveyed on funding models and funding levels that they're willing to support. DSS is projecting cost both for the current base services and for possible enhancements. All will be compiled into a report to go to the **USMAI's** Council of Library Directors, and to non-**USMAI** directors separately, along with the recommendation of a five year ongoing pricing plan.

Finally, with additional libraries wishing to join MD-SOAR, decisions need to be made about whether to allow this, and how to go about it, particularly in regard to a potential one-time fee to cover start-up costs. Adding new libraries may serve as a means of obtaining additional funds to pay for enhancements while keeping the price affordable for all.

Conclusions

With a substantive investment of time by a core group of leaders and experts from a handful of libraries, implementing a shared repository was challenging, yet successful.

continued on page 62

Optimizing Library Services — The OPAC

by **Edward Iglesias** (Web Services Librarian, 204 Mitchell Street, Nacogdoches, TX 75965) <edwardiglesias@gmail.com>

Column Editors: **Elizabeth Leber** (Promotions Assistant, IGI Global) <eleber@igi-global.com>

and **Lindsay Johnston** (Managing Director, IGI Global) <ljohnston@igi-global.com>

Column Editor's Note: *Promotions Assistant,* ***Elizabeth Leber****, joined the* ***IGI Global*** *team in November 2016, and she recently became a column editor for* ***Against the Grain****.* ***Elizabeth*** *earned her BA in English with a focus on secondary education from* ***Penn State University****. She then continued to earn a Master of Arts in Education: Adult Education and Training degree from the* ***University of Phoenix****. Her professional background was primarily focused on enrollment in higher education prior to transitioning to a marketing career in the publishing sector.* ***Elizabeth*** *currently resides in Palmyra, Pennsylvania. Most importantly, she is eager to collaborate with the outstanding* ***Against the Grain*** *team for* ***IGI Global's "Optimizing Library Services"*** *column, which focuses on what services academic libraries can offer in the 21st century. — **LJ***

When attempting to understand the way libraries acquire technology it is important to keep in mind that there was a time when nearly all technology was produced in house. The helpful Wikipedia article on OPACs ("Online Public Access Catalog." Wikipedia, the Free Encyclopedia, February 10, 2016. *https://en.wikipedia.org/w/index.php?title=Online_public_access_catalog&oldid=704231767*) gives a start time to online catalogs around 1975 with in-house systems developed at the **Ohio State University**. These were all in-house, locally developed systems since there were no ILS vendors until the 1980s. The records that went into those systems were developed largely by the **Library of Congress** in the 1960s ("MARC." Accessed April 5, 2016. *http://lili.org/forlibs/ce/able/course8/04marchistory.htm*). The earliest mention of the word OPAC is from around 1976 with **OCLC** (a library consortium that later became a library vendor) developing the first shared online catalog to be widely used. Throughout the 20th century, the technology of libraries was very DIY. Around 1980, all of this changed with the advent of cheap computing and vendors that offered products to libraries that previously only had card catalogs. Since then, more and more library technology has been purchased as a product from a vendor rather than being developed as a solution by staff.

Typically the transition from an in-house system to an outsourced system has a specific process: (1) there are cards that are typed up locally; (2) eventually this gets outsourced and cards are bought; (3) this information gets put into a database and is made available electronically; (4) the online catalog eventually replaces the print card catalog; (5) librarians who adopted the new platform became experts at searching the in-house system; (6) the vendor supported system takes its place; and (7) the in-house system is eventually retired. The vendor system is not as customizable as the old system, but everyone learns to make do. These precipitous declines in technology investment, customizability and local control are the hallmarks of outsourcing and will be seen again and again. As **Marshall Breeding** reported in 2007:

> "New Product Offerings from SirsiDynix" — SirsiDynix Symphony incorporates open, industry-standard technologies, offering the library community features and capabilities including: a service-oriented architecture (SOA), software-as-a-service (SaaS) options, power library "user experience" portal and search solutions, comprehensive integrated library management and productivity solutions, Java-based staff clients for all modules, fully documented application programming interfaces (APIs), Unicode support, advanced business intelligence and reporting tools, support for SIP2 and NCIP and support for the Oracle relational database management system. ("New Product Offerings from SirsiDynix: SirsiDynix Introduces SirsiDynix Symphony as New Integrated Library System." *Library Hi Tech News* 24, no. 7 (August 2007): 37–37.)

If this is the state of the art for OPACs, it is helpful to contrast what is gained and lost. After the first breed of home grown OPACs, the next generation focused on institutions that would largely maintain their own servers and network architecture. MARC records were loaded locally and were stored on the server. These records were very similar and had the same access points (author, title and keyword). Because MARC was designed at a time when memory was very limited, these records were stored in a flat file rather than a relational database. In order to search these records, there were indexes created at each of the access points. These records were stored on a system usually designed by information technology specialists at the institution. All of this meant that while the library had access to its own hardware and software, once a vendor became involved, the control was increasingly out of their hands. The migration from one OPAC to another requires the vendor's involvement because it was no longer a matter of just moving records. They had to be exported with customizations, which may or may not have been supported by the new system.

A hopeful change to this status quo is the growth of open source systems, which allows much more flexibility and local control. The tradeoff is the necessity for local expertise, specifically, in house programmers and systems administrators who are comfortable working with documentation and informal online communities as opposed to calling a help desk. As vendor support costs continue to rise, and the number of experts in open source systems grow, products such as Koha or Evergreen — especially when supported by independent companies such as **Bywater Solutions** — become much more realistic.

As OPACs became the de-facto inventory control system for libraries, many item types were hammered into place that were never meant to be supported. Dublin Core records imported from image or document repositories, were the first candidates. However, the real struggle came as electronic serials grew in prominence. Library systems and librarians had a great deal of expertise in dealing with paper serials. With the rise of online database aggregators, content became siloed into various database platforms. This prompted the need for a tool that would enable users to more easily find and retrieve content, and it would allow users to search across the entire library collection. Thus, was born the Discovery Layer.

Biz of Acq
from page 61

Both real dollar costs and the staff time investment were a fraction of what would have been needed to go it alone. Roadblocks came in the form of issues on which no consensus could be reached, and compromises that failed to satisfy any given campus but that served the overall needs of the platform and its users. For participating libraries, MD-SOAR jump-started repository programs that were lagging due to a lack of funding or staff time by substantively reducing those costs and technical competencies required of any single partner. During the pilot, the platform was successfully launched and policies developed to ensure an appropriate level of consistent usage of the platform by partners, allowing all more time to spend promoting their repository. Together we were readily able to do what all of us were struggling to do alone, and to do it better than any one of us might have done it alone.