

2014

Bridging the Gap between Digital Measures and Digital Commons in support of open access: Or, how I learned to stop worrying and love human mediation

Andrew Wesolek

Clemson University, awesole@clemson.edu

Follow this and additional works at: http://tigerprints.clemson.edu/lib_pubs

 Part of the [Library and Information Science Commons](#)

Recommended Citation

Please use publisher's recommended citation

This Article is brought to you for free and open access by the University Libraries at TigerPrints. It has been accepted for inclusion in Publications by an authorized administrator of TigerPrints. For more information, please contact awesole@clemson.edu.

TECHCAST

**Bridging the Gap between Digital Measures and Digital Commons in
Support of Open Access:
Or, How I learned to stop worrying and love human mediation**

ANDREW WESOLEK

Clemson University Libraries¹

Clemson, South Carolina

Introduction

Utah State University is home to a thriving institutional repository (called DigitalCommons@USU), an instance of the activity-reporting tool, Digital Measures (DM), and, as of May 30th, 2012, an institutional open access policy. In 2008, shortly after USU Libraries launched the institutional repository, efforts were made to harness the university-wide reporting capabilities of Digital Measures. The latter, managed by USU's Office of Analysis, Assessment, and Accreditation, was seen to be an important tool in developing the IR, as faculty are required to record their activities and intellectual contributions to this internal database. These initial efforts, however, were not largely successful and subsequently deprioritized.

A changing campus culture, the passage and implementation of an open access policy, and changes in the capabilities of Digital Measures, though, prompted

¹ This article describes work performed while the author served as Scholarly Communication Librarian at Utah State University. He held this position until August of 2013.

us to reexamine the relationship between Digital Measures and our institutional repository (IR) in hopes of finding new ways to automate the migration of data from the former to the latter. We found that while near-total automation appeared possible, in the interest of remaining in compliance with copyright law and ensuring the best quality metadata in our repository, it was perhaps not desirable. This article describes our initial attempts at DM and DigitalCommons @USU (DC) integration, the passage and implementation of USU's open access policy, and how this policy prompted a multi-faceted implementation approach that included a human-mediated harvest of material from DM.

Background and Initial Attempts at Automation

Committed to serve the public through learning, discovery, and engagement, Utah State University is the land and space grant institution of the state of Utah. Its locations across the state support nearly 29,000 students. 17,000 of these attend USU's main campus, which is located about 70 miles north of Salt Lake City, in Logan Utah. With a Carnegie classification of "high research," USU emphasizes engineering and education.

As part of its commitment to the public good and advancement of knowledge, USU invests heavily in scholarly communications initiatives. We were among the early subscribers to the bepress DigitalCommons platform, launching DigitalCommons@USU in November of 2008. This repository aims to capture, preserve, and provide open access to the scholarly output of the University.

The repository initiative has been largely successful. To date, the repository contains about 38,000 records. Full-text downloads, when available, total more than

two million, with about one million of these occurring within the past year. Moreover, our faculty participation rate, as measured by the number of USU-affiliated Selected Works pages stands at about 20 percent. The repository is managed by a scholarly communication librarian and between two to four student employees.

A key component of the success of our repository is the full-service approach taken by the library. The Merrill-Cazier library has a strong history of collaboration, both among library departments and between the library and other colleges and departments on campus. USU's first institutional repository manager, Heather Leary, capitalized on this collaborative spirit to engage subject librarians and inspire them to actively build partnerships in support of recruiting content for the IR. Subject librarians conducted a robust marketing and outreach campaign, and as a means to reduce administrative burdens on faculty, library staff took on the deposition of materials on behalf of USU's faculty (Leary, 2012).

In addition, subject librarians were also encouraged to perform copyright clearance on behalf of faculty. At USU, faculty submit their CV to their subject librarians, who then relay it to the IR staff. Students upload all bibliographic information to the IR and to a Filemaker Pro database that tracks permissions. Subject librarians then perform copyright clearance on each article, request manuscripts from faculty when permissible, upload manuscripts, and perform quality control on metadata entered in the IR.

This process had two drawbacks. First, it shifts a very substantial administrative burden to the subject librarians. As the repository grew, it became

necessary to shift this burden of copyright clearance and quality control from subject librarians to student employees (Wesolek and Lundstrom, 2012), for better or for worse. Second, the ingestion process is still dependent on faculty submissions to their subject librarians. While the model outlined above largely eased faculty concerns about potential administrative burdens involved with IR participation, it still depended on faculty alerting their subject librarians of new publications to be added to the IR.

In 2009, Leary and Programmer Analyst Brett Jones investigated the possibility of exporting data from USU's instance of Digital Measures to reduce administrative burdens on both faculty and librarians while providing IR staff with up to date information on faculty publications. At USU, uploading bibliographic information for one's intellectual contributions to DM is mandated by the Provost's office and tied to tenure and promotion evaluations. This being the case, we would expect DM to serve as a current and comprehensive source of potential metadata for the repository.

To accomplish this, Jones used ColdFusion to query the Digital Measures API. Cold fusion can walk XML easily, and Jones did this to generate excel spreadsheets that, with a bit of manual tweaking, matched bepress's batch ingest. At this time, Digital Measures did not offer to store files, and so this approach could only capture the metadata entered by faculty.

The drawbacks of this first attempt to harvest from Digital Measures to Digital Commons were three-fold. First, even though USU ties Digital Measures to promotion and tenure evaluations, faculty recording of their activities was spotty.

Some individuals appeared to update DM frequently, while others did so once per year, or just prior to any review period. It turned out, then, that DM was not an ideal way to keep current with faculty publications.

Second, our experience working with faculty CVs revealed startling inaccuracies in metadata. Quality control efforts on the part of IR staff, including checking CV entries against publisher's website's ensured quality metadata in the repository. However, the metadata faculty uploaded to DM was just as, if not more suspect than the metadata on their CVs. So, any automation achieved by harvesting from DM did not alleviate the staff time devoted to quality control.

Third, this effort did not address permissions. Faculty did not grant IR staff permission to allow the works that they had added to DM to be added to the IR. In addition, we were still faced with the challenge of copyright permissions. By mid-2011, USU's copyright clearance workflow had become unsustainable and the DM procedure did not have the potential to alleviate the administrative burden on IR staff and subject librarians.

So, while the potential to harvest metadata from DM in spreadsheets that could be ingested by USU's instance of Digital Commons was initially promising, the potential to streamline workflows was not there. As a result, by mid-2011, the effort was largely deprioritized and IR staff shifted efforts to emphasize conferences, grey literature, library-based publishing activities, and other sources of content that largely circumvented the copyright clearance workflow.

Open Access Policy Passage and Implementation

Also in late 2011, the Merrill-Cazier library began working with university stakeholders on the passage of an institutional open access policy. With an understanding that the passage and success of an OA policy depended largely on faculty buy-in, discussion began in the faculty senate (Shieber and Suber, 2013). The Senate then tasked the Library with the development of a policy, which proposed the adoption of a Harvard-style rights-based OA policy (Shieber, 2010). Months of discussion led to unanimous support for the policy by the faculty senate, which was then signed into effect by USU President Stan Albrecht on May 30th of 2012 (Utah State University, 2012).

The passage of USU Policy 535: “Open Access to Scholarly Articles” positioned USU among the leaders of public institutions with adopted open access policies. As such, there was not yet a clear pathway forward for effective implementation. While working on passage of the policy, USU did join the Coalition of Open Access Policy Institutions (COAPI), which offered a valuable venue for institutions with OA policies to share experiences and best practices for policy development and implementation. Other institutions were indeed further along the pathway than USU, but OA policy implementation was still largely experimental and based largely on individual institutional culture. Duranceau and Kriegsman (2013) have since published an excellent work on the implementation efforts of some of the first OA policy institutions.

Based on the experiences of some COAPI colleagues and the USU experience of developing its own IR, we determined that effective implementation of the OA policy would require the wide-spread delivery of a clear message, full service

deposition on the part of the library, and the ability to capture manuscripts shortly after publication. As Columbia's Rebecca Kennison would later put it, faculty are "fine with their stuff going into the repository, but they don't want to have to *do* anything (or at least as little as possible) to make that happen" (Duranceau and Kriegsman, 2013, p. 83). So, we needed to make faculty aware of the policy and repository, offer to *do everything* for them, and capture manuscripts before they are lost or discarded. This would require a multifaceted implementation approach.

On passage of the policy, Dean of Libraries Richard Clement charged Scholarly Communication Librarian, Andrew Wesolek with chairing an OA Policy Implementation Task Force, which would be composed of Wesolek and four subject librarians. The first priority of the task force was to craft a message to be delivered as a roadshow to the departmental retreats that occur prior to each Fall semester.

Crucial to crafting an effective message to faculty was a thorough understanding of which arguments and terminology would resonate with faculty and which would not. Our early discussions with the Faculty Senate illuminated the fact that without a thorough understanding of the purpose and mechanisms of OA policies, faculty can perceive them to be a threat to their ownership of their scholarly works. With faculty buy-in a critical component for the success of the policy, then, it was extremely important to develop a concise message that would resonate with faculty.

To accomplish this, the Task Force called together a focus group to take part in the development of the roadshow. To ensure our message resonated across departmental boundaries, the focus group was composed of representatives from

the hard sciences, applied science, humanities and social sciences, and one member of administration. In addition, focus group participants consisted of known OA supporters, opponents, and several “undecideds.”

This group met with the task force three times through the summer of 2012 as the task force honed its message. We found that at USU impressions were most positive when the message focused on author’s rights, the benefits of OA to researchers, and individual cases where OA research provided substantial benefits to researchers. Arguments from USU’s status as a public land grant, benefits to the public, and efficient use of public research investments did not resonate with our faculty. Once our message was crafted, we developed a roadshow presentation that we could deliver to colleges and departments on campus (Scholarly Communications Office, 2012).

Once again relying on a robust marketing campaign undertaken by our subject librarians, the OA task force secured time on greater than 90% of departmental retreat agendas. This was a tremendous success for our outreach efforts. However, we soon found that we relied too heavily on faculty to initiate the deposit process. Our workflow requested faculty to simply email their manuscripts to a generic Scholarly Communications Office address or write to request a waiver. Despite our outreach efforts, the end of our first semester of OA implementation saw only six manuscripts added to the IR as a direct result of the policy.

Clearly we needed to adjust our approach to make compliance with the new policy less burdensome on faculty. IR staff began developing a two-fold approach, which would both notify staff when works were published, but also “meet faculty

where they are” by reinvestigating the use of Digital Measures. First, the Scholarly Communication Librarian set up alerts through Scopus to notify him when USU faculty published new works. These would come in a weekly digest and prompt IR staff to contact faculty directly to request manuscripts for deposit in the IR pursuant to the OA policy. This approach was moderately successful, but also extremely labor intensive. Typically, connections made with faculty resulted first in the submission of a publisher’s PDF of the work. IR staff used this as an opportunity to educate faculty on the difference between a publisher’s version and a final accepted manuscript of a scholarly article. In greater than 50% of instances, this contact and education effort led to the deposit of the appropriate version of the work in the IR.

At the conclusion of the second semester of policy implementation, contacting faculty shortly after publication resulted in the deposit of 47 additional manuscripts. However, with limited staffing support, it was not possible for the Scholarly Communication Librarian to contact each faculty author individually. While the percentage of faculty contacts that led to a deposit was quite good, the percentage of faculty contacted initially was quite low. So, while contacting faculty in response to Scopus alerts was a valuable tool for cultivating awareness of current publications in support of policy implementation, the labor involved indicated that it could be only one facet of a multi-faceted implementation strategy.

Reevaluation: Technology and Workflow

The limited success of the initial implementation strategy focused on faculty email submissions prompted us to adopt a multifaceted approach. The Scopus alerts described above were labor-intensive and difficult to maintain given our staffing

constraints. Recognizing the need to further reduce administrative burdens on faculty and capture manuscripts before they are discarded, the Scholarly Communication Librarian reconnected with Programmer Analyst Jones to reopen the investigation into linking Digital Measures and Digital Commons.

In the time since the initial attempt had been deprioritized, DM now offered hosted storage space. This opened a great opportunity for OA implementation: faculty were required to update DM. If we could prompt them to deposit their manuscripts in DM at the time of updating, we would have a chance at capturing and depositing them before they were discarded while avoiding the investment of human resources needed to contact each individual.

First, we opened a ticket with DM to include two additional non-required metadata fields where faculty could upload final accepted manuscripts and publisher PDFs (see Figure 1).

[INSERT FIGURE 1 HERE]

Our experience contacting faculty individually informed us that a single field requesting a “final accepted manuscript” would result in the uploading of an overwhelming majority of publisher PDFs. Our hope was that by adding a second field, called “Upload published version of record,” we could eliminate some of the confusion. In addition, we make explicit that these manuscripts are to be deposited

in the Digital Commons, and so, we take uploading a manuscript to be a tacit granting of permission for us to deposit their works.

Jones then offered the use of a 500 gigabyte WebDAV enabled server—for comparison, our IR is only about 114 gig total—for this effort and opened a ticket with DM to mirror to his server. DM employs rather strict security protocols, but by mirroring content to the WebDAV server, the Scholarly Communication Librarian could access and download content. We then requested DM create a report that contained the necessary metadata for the IR and captures only those submissions that contained a file in the “upload final accepted manuscript field.” These would appear as a URL in the DM spreadsheet that would link to the file on our in-house WebDAV server. See Appendix 1 for an example of this report.

By the middle of the Fall 2012 semester, we had established the server, new metadata fields in DM, and the ability to generate the appropriate report. We planned to conduct a quarterly harvest of DM as a means to supplement manuscripts obtained via direct faculty outreach. We also began advertising this feature to faculty as an easier, albeit slower (due to the quarterly harvest) way to comply with the OA Policy. To allow time for outreach efforts to take effect, we determined to implement our first harvest of DM in June of 2013. Activity on the WebDAV server did indicate that faculty began depositing work before any marketing attempts were made, though—illustrating a demand for this type of service.

To harvest from DM and upload to the IR, the Scholarly Communication Librarian runs a report in DM to capture all of the items uploaded in the “final

accepted manuscript” field within a specified date range. A student employee then creates a column between “G” and “H” (see appendix 1) and labels it “Sherpa Permissions.” The student then checks permissions for each article; indicating when the deposition of acceptable manuscripts is appropriate and if any embargo periods apply.² When permissions are appropriate, the student downloads the file from the WebDAV server, stores it on a local server, then indicates the filename in the “W” column. The Scholarly Communication Librarian then checks each file to ensure the version is appropriate and provides a list of appropriate uploads to the student.

This is obviously not a highly automated process. As we moved forward, we realized that human mediation is an essential component of this process to ensure quality metadata in the IR and to avoid copyright infringement. First, initial contact with faculty uncovered a lack of understanding regarding the version of manuscript appropriate for deposit in the IR and we were concerned—rightly, as evidenced by the results of our first harvest—that any fully automated approach would result in a significant number of publisher PDF’s uploaded to the repository. Second, previous IR experience taught us the necessity of quality control for the metadata uploaded from faculty CVs. Going into this process, we were concerned that metadata uploaded to DM would require similar, if not more, quality control efforts.

Results and Conclusion

The initial harvest of DM was quite successful, but illustrated the necessity of additional efforts to educate our faculty while justifying our caution in the pursuit of

² I should mention here that USU made the strategic decision not to take a robust stance on the license to distribute these manuscripts in our IR until the policy was well established and faculty support was significant. As such, we still check Sherpa permissions for each article.

full automation. 415 articles were deposited in DM prior to the first harvest, but only 75 of these were both the appropriate version and had been published in journals that allowed us to disseminate them through the IR. A breakdown of submissions and acceptance by college may be found in Table 1.

[INSERT TABLE 1 HERE]

While we could only upload 18% of the articles submitted, it is important to point out that if this process was fully automated, we could have potentially had several hundred copyright infringements on our hands. In addition, continued education of faculty, both on an individual basis through Scopus-prompted contact and in group instructional settings, is expected to increase this percentage over time. The more widespread the understanding of the distinction between versions, the greater percentage of submitted material can be uploaded to the IR.

So, in our experience there is no single “right way” to implement an OA policy. As noted earlier, approaches will likely be influenced by institutional culture and entrenched idiosyncrasies. In the case of USU, a multifaceted approach is proving most successful; one that incorporates a human-mediated harvest of DM as one component. While this human-mediated harvest of metadata and files from DM suffered similar setbacks to our first attempt—metadata is not always pristine, and content is not always current-- when coupled with individual outreach prompted by Scopus, and additional education and outreach efforts, it is an important component of a multi-faceted approach to OA Policy implementation.

REFERENCES

- Duranceau, Ellen Finnie, and Sue Kriegsman. 2013. *Implementing Open Access Policies Using Institutional Repositories*. American Library Association. Association for Library Collections and Technical Services.
<http://hdl.handle.net/1721.1/76721>
- Leary, Heather, Kacy Lundstrom, and Pamela Martin. 2012. "Copyright Solutions for Institutional Repositories: A Collaboration with Subject Librarians." *Journal of Library Innovation* 3(1): 62-71.
- Scholarly Communications Office. 2012. *USU Policy 535: Open Access to Scholarly Articles. What it is and what it means for us*. Last modified January 2012.
http://works.bepress.com/andrew_wesolek/5
- Shieber, Stuart and Peter Suber. 2013. *Good Practices for University Open-Access Policies*. Harvard Open Access Project. Last modified: September 15, 2013.
bit.ly/goodoa

Shieber, Stuart M. 2010. "Model Open Access Policy." *Harvard Office for Scholarly Communication*. Retrieved September 15, 2013 from:
https://osc.hul.harvard.edu/sites/default/files/model-policy-annotated_01_2013.pdf.

Utah State University. 2012. *Open Access to Scholarly Articles (535)*. Utah State University Policy Manual. Last modified May, 2012.
<http://www.usu.edu/hr/files/uploads/Policies/535.pdf>

Wesolek, Andrew and Kacy Lundstrom. 2012. "How subject librarians make better IRs and how IRs make better subject librarians." Presented at Utah Library Association, April 2012. http://works.bepress.com/andrew_wesolek/1/