



Volume 5, General Issue (2017)

Workflow Development for an Institutional Repository in an Emerging Research Institution

Jeanne Hazzard, Stephanie Towery

Hazzard, J. & Towery, S. (2017). Workflow Development for an Institutional Repository in an Emerging Research Institution. *Journal of Librarianship and Scholarly Communication*, 5(General Issue), eP2166. <https://doi.org/10.7710/2162-3309.2166>



© 2017 Hazzard & Towery. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

Workflow Development for an Institutional Repository in an Emerging Research Institution

Jeanne Hazzard

Taxonomy Manager, MissionBox

Stephanie Towery

Copyright Officer, Texas State University

INTRODUCTION This paper describes the process librarians in the Albert B. Alkek Library at Texas State University undertook to increase the amount of faculty publications in their institutional repository, known as the Digital Collections. **DESCRIPTION OF PROGRAM** Digital Collections at Texas State University is built on a DSpace platform and serves as the location for electronic theses and dissertations, faculty publications, and other digital Texas State University materials. Despite having launched the service in 2005, the amount of faculty work added to the repository has never been at the levels initially hoped for on launch. **DEVELOPMENT AND IMPLEMENTATION OF THE WORKFLOW** Taking a proactive and cooperative approach, a team of librarians developed and piloted a workflow, in which library staff would retain the already established protocol of gaining faculty permissions prior to uploading material while respecting publisher copyright policies. **RESULTS** Prior to the vita project, the repository archived 305 faculty publications total. Fifty-seven were added during the pilot, which represents an 18.5% increase. Of a total of 496 articles, seventeen titles were found in the blue category, which allows publisher pdfs to be archived. The majority of articles (233) were found in the green category, which allows either a pre- or a post-print copy of an article to be archived. One hundred ten of the identified titles were in the yellow and white journal categories, representing 22% of our total, and the team was able to archive only five of these. Finally, 16% (81) were not found in the SHERPA/RoMEO database (color-coded beige). Only 18 of these articles were archived. **ASSESSMENT** We discovered that our faculty retain nearly none of their pre-print or post-print versions of their published articles, and so we are unable to archive those titles in the repository. Nearly 47% of the articles found were in green journals that allow only pre- or post-print copies. Most faculty were unable to produce versions of their work other than the publisher's PDF, which many publishers restrict from upload into a repository.

Received: 11/07/2016 Accepted: 05/30/2017

Correspondence: Stephanie Towery, 601 University Drive, Alkek 350D, San Marcos, TX 78666, sst25@txstate.edu



© 2017 Hazzard & Towery. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

INTRODUCTION

This paper traces the development and implementation of a workflow intended to increase the number of faculty scholarly articles in the institutional repository at Texas State University. Founded in 1899 in San Marcos, Texas, Texas State University has a student population of over 38,000 students and offers 90 masters and 12 doctoral programs. In 2012, Texas State University was reclassified as an Emerging Research University by the Texas Higher Education Coordinating Board and is the fourth largest public university in Texas and the largest university of the eight universities in the Texas State University System. The Albert B. Alkek Library serves Texas State University's 38,000 students and 1,300 full-time faculty. Texas State University was originally chartered as a small teacher preparation institution. And as a teaching institution, faculty time has been devoted largely to instruction, with much less emphasis on research. Yet attaining Emerging Research status means Texas State University's scholarly communication needs are in transition to a new, more research-focused environment.

In 2004, librarians at Albert B. Alkek Library began discussing plans for opening an institutional repository to house and promote both faculty publications and electronic theses and dissertations produced by the university community. As early as 2002, the academic library community was promoting the development of institutional repositories as a solution for problems, costs, and barriers via traditional publishing models. Crow (2002) stated "[i]nstitutional repositories represent the logical convergence of faculty-driven self-archiving initiatives, library dissatisfaction with the monopolistic effects of the traditional and still-pervasive journal publishing system, and the availability of digital networks and publishing technologies" (p. 29). By late 2005, the library implemented an institutional repository and began accepting faculty self-submissions. Library leadership created a new librarian position, Digital Collections Repository Librarian, which would oversee the administration, function, and design of the institutional repository.

Faculty uptake in the repository service was low, which was not uncommon in other academic repositories. Despite the best of intentions for providing a new open-access model of academic publishing, institutional repositories have not been able to convert an entrenched model of scholarly output to one of an institution-based service. Chan (2004) notes a similar low rate of participation at the University of Toronto, citing,

cultural inertia is often cited by faculty members as the reason for the slow adoption of self-archiving. Lack of awareness of the importance of open access is another common reason. Lack of trust in institutional commitment to the long-term maintenance of the repository could also be a factor (p. 293).

Despite the unenthusiastic faculty participation levels, library leadership still found value in the repository, supported by the amount of downloads of repository content. Since its 2005 launch, the number of total downloads, (3,204,183), with an average annual increase in downloads of thirty percent over the previous seven years, demonstrated that even though the repository contained mostly theses and dissertations, it still proved to be a useful tool to promote the research and scholarship produced by the university.

Library leadership and staff proposed that increasing the amount of faculty publications in the repository could be achieved with a new strategy. In early 2014, library leadership created the Scholarly Communication Team, charged with raising awareness and fostering understanding about scholarly communication issues and trends to the Texas State University campus community. Initially, the team was composed of the Head of Research, Instruction, and Outreach (Team Chair); two other Research, Instruction, and Outreach Librarians; the Copyright Officer; the Collection Development Librarian; and the Library System Coordinator. Garnering more faculty publications into the repository is another of the charter goals of the Scholarly Communication Team.

Library staff at the Albert B. Alkek Library saw the importance of the repository as a promotional tool for the university and the scholarship it produces, and the team hoped to advance its mission by adding faculty research previously published in open access archiving-friendly journals. A new Copyright Officer joined the team in 2015, as did the library's Digital Collections Repository Librarian, who designed an initial workflow that became the current workflow after a pilot and review by the Scholarly Communications Team. The pilot consisted largely of the Copyright Officer and the Digital Collections Repository Librarian working together to move faculty publications through the workflow and into the repository.

The team did not establish explicit success conditions for the pilot, having experienced years of self-submissions lower than library leadership had anticipated. Despite some skepticism from some team members, we believed that adding librarian facilitation would increase numbers significantly. The team's expectations were low, but the costs were also low and consisted mainly of the staff time of two full-time librarians. The two librarians believed at the outset that the pilot would take a few hours a week of their time.

LITERATURE REVIEW

Since the beginning of institutional repository development in academia, administrators have been making efforts to promote the service as valuable for scholarly publishing and open access, and then trying to discover why faculty uptake of the service is not greater

than anticipated. A review of the literature around the development and implementation of repositories shows a general focus on areas such as awareness and marketing of repository services, including perceptions and reactions of intended user groups, copyright issues, and workflows.

The literature reveals that faculty reluctance to submit to institutional repositories is widespread. Even when an institutional mandate requires deposit of articles to a repository, faculty may not necessarily follow through, as library staff discovered at Oregon State University. Zhang et al (2015) note that “the expectation was that the approval of the policy would increase faculty motivation to deposit articles and expand OSULP’s ability to request manuscripts,” but “passing an OA [open access] policy alone is not a guarantee of increased faculty engagement in OA initiatives” (p. 9).

In fact, it may be that open access mandates may have the opposite intended effect of increasing institutional content into repositories. In 2014, Texas A&M University conducted a survey on faculty awareness and perceptions of the institutional repository. Yang and Li (2015) discovered that while there was a general sense of awareness at a relatively high level—90% of faculty respondents were aware of open access journals—far less held a positive attitude towards mandated publishing in open access journals or repositories (p. 12).

Only a little over half of the respondents agree that if TAMU adopts OA mandates, their work will be read by more people and will reach more people outside of their fields. They are highly skeptical as to whether OA mandates will help them secure grant funding, and do not believe a mandate would be easily complied with (Yang and Li, 2015, p. 13).

Alternative approaches have had different outcomes. Ferreira, et al. (2008) had a great deal of success increasing faculty deposit by combining a mandate with financial incentive. The University of Minho contributed a significant financial incentive towards their repository project. For the first two years after the mandate, faculty departments would receive money whenever faculty deposited work in the repository. With this combination of mandate and incentive, the proponents of the repository were able to significantly increase faculty input.

University of Minnesota librarians decentralized their scholarly communications efforts in part by making departmental liaisons responsible for assisting in the recruitment of faculty work for the repository (Malenfant, 2010). Prior to soliciting faculty for publications for their repository, the University of Minnesota libraries instituted a strategic change

to “define baseline expertise in scholarly communication for all librarians who serve as liaisons to disciplinary faculty members” (p. 64). The University of Minnesota spread the responsibility for scholarly communications goals among the liaison librarians, so they were personally invested in the success of scholarly communication goals, such as soliciting faculty for publications (Malenfant, 2010, p. 69).

Regardless, getting faculty to post their publications in an institutional repository has always been difficult. Mercer, Rosenblum, and Emmet (2007) note that “persuading faculty to fill institutional repositories (IRs) through self-archiving remains challenging” (p. 190). Changing faculty minds on desirable publishing platforms is equally difficult. Confusion regarding copyright, intellectual property rights, and publishing agreements also plays a role in the lack of participation in institutional repositories. In a study of barriers to institutional repository participation, Kim (2010) found, among other things, “two factors were found to impede self-archiving: concerns about copyright and additional time and effort” that active participation in repository publishing requires (p. 1920). Suggestions for easing faculty concerns and workload include offering more information, workshops, and assistance with the copyright clearance process. Leary, Lundstrom, and Martin (2012) found that,

[t]he copyright clearance process involves many steps but follows a simple pattern of logic, beginning with identifying who the copyright owner is and what permissions they allow for the work. It becomes more complicated as copyright owners sometimes do not allow using a specific version of a published work in an IR. Working through this process has the potential to be time consuming and requires direct contact with the publisher (p. 104).

Addressing the time commitment, Kim (2010) asserts that “technical and logistical assistance for self-archiving would encourage faculty who are less adept at computers to participate,” and that “this support may also alleviate faculty concerns about the extra time and effort inherent in self-archiving” (p. 1920).

Still, hurdles relating to awareness remain. Indeed, a lack of awareness has been recognized as an ongoing issue with faculty self-archiving, in spite of the usual library marketing through newsletters, informational emails, and workshops:

Despite our best efforts to make faculty aware of the abundance of resources made available by the Libraries, it seems that our audience continues to remain unaware of some of our services and resources. This only reinforces the need for continuous communication (Yang and Li, 2015, p. 1).

It could be argued that libraries and administrators have not done a thorough job of marketing. Chan (2004) recognized awareness and clarity of purpose as a barrier to participation citing “lack of awareness of the importance of open access is another common reason” for lower participation rates (p. 293). The intent, purpose, and benefit of adding content to an institutional repository and of open access publishing have not been emphasized enough. In an assessment of repository services at Carnegie Mellon University, Covey (2011) discovered via focus groups that “[l]acking awareness, participants also lacked understanding. They asked many questions about scope, motivation, and operational details” (p. 9). These focus groups also revealed a concern of the time commitment of vetting the materials for copyright clearance before submitting:

[N]o one objected to the repository or to the Libraries harvesting work they had already self-archived, but many perceived manually harvesting that work and, going forward, expecting faculty to provide metadata and copies for deposit as too slow and labor intensive (Covey, 2011, p. 9).

DESCRIPTION OF PROGRAM

The Alkek Library’s repository staff chose to view and promote the repository to faculty as a service that could provide an access and discovery point to users who may not be directly affiliated with the university. By taking a service approach, rather than promoting the repository as a replacement for traditional publishing channels, library staff hoped to gain a higher rate of faculty acceptance and comfort with the repository as a distribution platform. Texas State University does not have an open access mandate, so library staff must rely on faculty to participate voluntarily.

In recognition of the many challenges of increasing faculty publications in its institutional repository, the team developed a pilot project that would address concerns about copyright clearance and the involved time commitment on behalf of faculty. The process required the repository administrator, copyright officer, and subject librarians to work collaboratively.

Development and Implementation of the Workflow

The library intended the Digital Collections repository to grow through deposit of electronic theses and dissertations and voluntary deposits of scholarly work by faculty authors. The repository allows faculty to self-submit, and the library encouraged faculty to take advantage of the self-submit function to increase the reach of their scholarly work. While a few individuals were prolific users of the self-submit function, the majority of publishing

faculty did not self-submit or ask library staff to assist them in uploading their publications.

Library leadership tasked the library's Scholarly Communications Team with several strategic plan goals related to faculty outreach and open access. One of the goals of the team was to facilitate the deposit of more scholarly material to the Digital Collections repository. Workflow development was driven by tools at hand and established relationships: SHERPA/RoMEO and the subject librarians' faculty contacts in the different departments and colleges. Subject librarians contacted faculty about their willingness to send the library their vitae. If subject librarians' efforts were successful, all the copyright vetting, acquiring publisher permissions for published works, and deposit of the publications would be handled within the library.

The initial workflow was created by the Digital Collection Repository Librarian and the Copyright Officer, and relied on the look-up function in SHERPA/RoMEO. SHERPA, which stands for Securing a Hybrid Environment for Research Preservation and Access, supports a service which lists publishers' self-archiving policies by journal. RoMEO, currently run by SHERPA Services at the Centre for Research Communications, University of Nottingham, UK, was originally created as the RoMEO Project at the University of Loughborough, UK. RoMEO is a "searchable database of publisher's policies regarding the self-archiving of journal articles on the web and in Open Access repositories." (Millington, 2011, p. n.) RoMEO has proved to be an invaluable tool for the open access archiving process.

Starting with one faculty member's curriculum vita, the Copyright Officer and the Digital Collections Repository Librarian, who was the repository administrator, tested a potential workflow. Using SharePoint as a collaborative workspace, the Digital Collections Repository Librarian transcribed faculty publication data into a spreadsheet, sorted by journal title and referenced in SHERPA/RoMEO. The time devoted to looking up the journals in SHERPA/RoMEO varied greatly by length of CV. A twenty-page CV with numerous scholarly articles in a variety of journals could take several hours. The color categories of SHERPA/RoMEO indicate the publishers' policies toward open access archiving and simplified the sorting and categorizing of the different articles after transcribing. RoMEO uses four colors to categorize rights: blue, green, yellow, and white. The different colors represent different levels of publishers' willingness to support reproduction of articles in an open repository.

Many journal titles were not found or had no official designation in RoMEO, so we chose to color code those titles in beige. White represents journals that are listed in RoMEO but

that have not provided RoMEO with information about their open access archiving policies. Therefore, white and beige coded articles represented articles for which we had little to no information. We anticipated that these journals might have potential for allowing posting in the repository. With the transcription and color coding complete, the Copyright Officer prepared permission requests to white and beige publishers.

The Copyright Officer found contact information for white and beige coded publishers and requested permission to post the articles in the Digital Collections repository. Upon receiving the publisher replies, the Copyright Officer uploaded copies of the permission emails to the SharePoint folder. The beige journals were almost exclusively professional organization newsletter or magazines or very small publications not associated with a university. Generally, the small-scale nature of the beige publications made the permissions process more difficult.

Communication with the white and beige journals was by email to the editors, who each agreed to permit the publisher version of the article to be uploaded. Locating contact information for the beige journals and waiting for responses was the most time-intensive portion of the pilot. The Copyright Officer contacted each journal at least twice by email before abandoning attempts at communication. The Copyright Officer could not identify and locate contact information for some of the journals. For those journals, the Copyright Officer requested additional information from the submitting faculty, through the mediation of the Subject Librarian, but no faculty submitted further information.

With copyright clearance taken care of and reproducible copies identified, the Copyright Officer pulled the publisher PDFs for archiving. For beige journals, most of the articles were either not available online or available on the open web through organizational websites. The Copyright Officer pulled publisher PDFs for green- and blue-coded journals from library subscriptions. The Digital Collections Repository Librarian took over again, and uploaded the PDFs into the repository. Most of the communication between the Copyright Officer and the Digital Collections Repository Librarian occurred via email or in person.

A wider effort to reach faculty was then launched by the members of the Scholarly Communications Team, who were also subject librarians, by contacting their faculty to solicit interest in posting to the Digital Collections repository. Continuing with the efforts to leverage established collaborative relationships, the team thought that the subject librarians should remain the contact point for the solicitation of work for the repository. Traditionally at Alkek Library, departments communicate with library staff through the mediation of their liaisons.

Results

The subject librarians received twenty-eight vitae from faculty from a variety of disciplines, with the greatest responses, four each, from biology and health sciences. Other disciplines that volunteered vitae included communication disorders, history, physics, and political science. A total of 496 articles were entered into spreadsheets and sorted into SHERPA/RoMEO categories. Seventeen titles were found in the blue category, which is the most open access archiving–friendly category. The majority of articles (233) were found in the green category, which is open access archiving–friendly with a twist. Journal publishers with a green designation allow either a pre- or a post-print copy of an article to be archived in an institutional repository. Prior to the vita project, the repository archived 305 faculty publications total. Fifty-seven were added during the pilot, which represents an 18.5% increase in that category. One hundred ten of the identified titles were in the yellow and white journal categories, representing 22% of our total, and the team was able to archive just five of these. Finally, 16% (81) were not found in the SHERPA/RoMEO database (color-coded beige). Only 18 of these articles were able to be archived.

Assessment

Among our findings, we discovered that our faculty retain nearly none of their pre-print versions of their published articles, and so we are unable to archive those titles in the repository. Nearly 47% of the articles found were in green journals that allowed only pre- or post-print copies. Most faculty were unable to produce versions of their work (pre-prints) other than the publisher's PDF, which many publishers restrict from upload into a repository. One solution to this problem is to educate faculty to keep versions of their work in the future. Yet this approach poses problems because currently faculty do not have a document management system in which to store and track their research. Faculty at Texas State University store their research on computer drives, with unknown naming and organizational structures.

Another obstacle to this method of harvesting faculty work for open access archiving is mainly time. The process of transcribing publication data into spreadsheets is time consuming, particularly when there is no standard vita format. Every faculty vita is as unique as a fingerprint, and this can make the manual process onerous. Fitting the demands of the project into multiple librarians' schedules caused backlogs and bottlenecks, which we hope to eliminate in the future by using a new institution-wide credentialing method that streamlines and normalizes all faculty vita information.

The pilot relied on subject librarians to mediate communications between the two librarians working on the pilot and faculty members. Before the pilot began, the team created an email template that the subject librarians could send to the faculty in their assigned departments, and the team held a meeting with the subject librarians before the pilot to answer their questions about the process. The lack of faculty CVs submitted from some departments may be due to a reluctance among the subject librarians to solicit faculty for CVs. Corroborating the outcomes of the Minnesota report, the team saw greater success with faculty from departments to which subject librarian members of the team were assigned. Other factors may have contributed to the lack of participation from some departments at Texas State University. Such factors might include cultures within departments, differing attitudes about Open Access among faculty, differences in publishing norms in different disciplines.

NEXT STEPS

The Scholarly Communication Team views the results of the pilot as a success, considering the overall number of vitae that were submitted and the extent of content that we were able to archive. But the team recognizes that elements of the project can be streamlined, particularly by relying less on librarian and faculty schedules and priorities. Texas State University faculty are moving the information in their vitae to a campus-wide system that organizes and displays all CV data in the same way. From this system, the repository administrator will be able to generate reports of all faculty publication data directly into a CSV file. From the CSV file, sorting and vetting the publication information should be a simple process. The team would like to incorporate this process into the workflow. In addition, the team will encourage subject librarians to invest in the success of the project, for example by taking on the tasks of checking the journals in Sherpa/RoMeo and pulling the publisher PDFs for their departments.

As Texas State University advances to Research University status, we also see opportunity for more outreach, in the form of education and workshops, on the significance and value of retaining preprint copies of published articles and management of publishing agreements, in which open access archiving policies are more easily tracked. In early 2017, in response to the pilot, the team developed and presented to faculty several library guides and presentations to try and counter negative faculty impressions about Open Access. While new faculty are the obvious targets of outreach, we feel there is value in encouraging established and tenured faculty to also rethink preprint archiving and access.

REFERENCES

- Chan, L. (2004). Supporting and enhancing scholarship in the digital age: The role of open-access institutional repositories. *Canadian Journal of Communication*, 29(3). <https://doi.org/10.22230/cjc.2004v29n3a1455>
- Covey, D. T. (2011). Recruiting content for the institutional repository: The barriers exceed the benefits. *Journal of Digital Information*, 12(3).
- Crow, R. (2002). The Case for Institutional Repositories: A SPARC Position Paper. Washington DC: Scholarly Publishing & Academic Resources Coalition (SPARC). Available: https://ils.unc.edu/courses/2015_fall/inls700_001/Readings/Crow2002-CaseforInstitutionalRepositoriesSPARCPaper.pdf
- Ferreira, M., Rodrigues, E., Baptista, A. A., Saraiva, R (2008). Carrots and sticks: Some ideas on how to create a successful institutional repository. *D-Lib Magazine*, 14(1/2). <https://doi.org/10.1045/january2008-ferreira>
- Kim, J. (2010). Faculty self-archiving: Motivations and barriers. *Journal of the American Society for Information Science and Technology*, 61(9), 1909-1922. <https://doi.org/10.1002/asi.21336>
- Leary, H., Lundstrom, K., & Martin, P. (2012). Copyright solutions for institutional repositories: A collaboration with subject librarians. *Journal of Library Innovation*, 3(1), 101-110.
- Malenfant, K. (2010). Leading change in the system of scholarly communication: A case study of engaging liaison librarians for outreach to faculty. *College & Research Libraries*, 71(1), 63-76. <https://doi.org/10.5860/0710063>
- Mercer, H., Rosenblum, B., & Emmet, A. (2007). A multifaceted approach to promote a university repository. *OCLC Systems and Services: International Digital Library Perspectives*, 23(2), 190-203. <https://doi.org/10.1108/10650750710748496>
- Millington, P. (2011). SHERPA/RoMEO FAQ - Publisher copyright policies & self-archiving. Available at: <http://www.sherpa.ac.uk/romeo/faq.php#whatis>
- Yang, Z. Y. & Li, Y. (2015). University faculty awareness and attitudes towards open access publishing and the institutional repository: A case study. *Journal of Librarianship and Scholarly Communication*, 3(1), eP1210. <https://doi.org/10.7710/2162-3309.1210>
- Zhang, H., Boock, M., & Wirth, A. (2014). It takes more than a mandate: Factors that contribute to increased rates of article deposit to an institutional repository. *Journal of Librarianship and Scholarly Communication*, 3(1), eP1208. <https://doi.org/10.7710/2162-3309.1208>