

Content liberation!

How increasing the institutional repository content turned into faculty outreach services

Jennifer Solomon

University of North Carolina at Chapel Hill

jsolomon@unc.edu

Rebekah Kati

University of North Carolina at Chapel Hill

rkati@email.unc.edu

Author's note

We have contributed equally to this project. We would like to thank Stefanie R. Knauer for her editing expertise.

Abstract

In January 2016, the University of North Carolina at Chapel Hill faculty adopted the Open Access Policy, which encouraged faculty to deposit their articles into the Carolina Digital Repository. The UNC-Chapel Hill Libraries Open Access Implementation team was then charged with increasing the amount of content in the Carolina Digital Repository and raising faculty awareness of the Open Access Policy and author rights issues. In this paper, we will discuss the challenges of locating and harvesting content, the outreach strategies we used with faculty from diverse departments, and the assessment of the overall project's success. We also share findings from our analysis of the content we collected and recommendations for replicating or scaling up similar projects.

Keywords

- open access
- scholarly communications
- institutional repositories
- outreach
- behavior change
- deposit requirements
- author rights
- collaboration

Introduction

The University of North Carolina-Chapel Hill (UNC-Chapel Hill) Libraries manages its open access content using the Carolina Digital Repository. The Carolina Digital Repository is a custom repository built on Fedora, a popular open source repository system, and custom Java applications. It launched in 2009 primarily to house student papers and supported mediated deposit. UNC-Chapel Hill Libraries needed a place to store its born-digital special collections objects and elected to use the Carolina Digital Repository for this purpose. Over time, the majority of submissions to the Carolina Digital Repository were special collections materials. With the passage of the Open Access Policy in 2016, there was renewed interest in optimizing the repository for scholarly content.

As part of the Open Access Policy implementation plan, UNC-Chapel Hill Libraries hired two new positions in 2017: an open access librarian (Jennifer) and an institutional repository librarian (Rebekah). As the open access librarian, Jennifer leads the outreach and marketing strategy to raise awareness of the Open Access Policy and increase the accessibility of UNC-Chapel Hill faculty research. In her role as the institutional repository librarian, Rebekah manages the Carolina Digital Repository operations, including all deposit workflows and policies. To operationalize this plan, we work closely together and also in collaboration with, the scholarly communications officer and the repository program librarian, the two other members of the Open Access Implementation Team.

Although the Open Access Policy had been in effect for over a year, data from Carolina Digital Repository deposit rates and anecdotal information showed that the campus was largely unaware of both the Open Access Policy and the Carolina Digital Repository; therefore, we envisioned the Content Liberation project as the first phase of a multi-year plan, to allow us flexibility to learn from our process and revise our tactics and goals iteratively. We view open access as a suite of services that the Libraries provides for the UNC-Chapel Hill community; inclusive of all faculty, staff, and students. However, the Open Access Policy is specific to tenure-track faculty who are publishing journal articles since 2016. Accordingly, when we began our roles in October 2017, one of our first priorities was identify key

faculty and their publications that met the criteria of the Open Access Policy. UNC-Chapel Hill does not have a centralized tracking system for faculty publications, thus we launched the Content Liberation Project to pilot three new tactics for recruiting content to the Carolina Digital Repository. As a complementary benefit, the process of recruiting content was an opportunity for outreach to both individual faculty and departments across campus.

Project Goals and Considerations

Our goal for the Content Liberation Project was to identify current, paywalled, and high impact content to be ingested into the Carolina Digital Repository. We considered multiple strategies to accomplish this, but realized that a one-size-fits-all approach was not feasible due to the large amount of UNC-Chapel Hill-authored scholarly output and variety of subject areas represented. The considerations which affected our approach were the presence of affiliations, depositors' familiarity with open access, the type of content deposited and the goals of our team.

Consideration 1: Presence of Affiliations

Due to the large number of papers published by UNC researchers, we wanted to automate the collection process as much as possible. Affiliations appeared to be a good, programmatic way to identify UNC-Chapel Hill authored content, since affiliations link the article's author with UNC-Chapel Hill explicitly. However, after collecting an initial batch of UNC-Chapel Hill-authored articles, we realized that this approach would return more articles in science, technology and mathematics fields, (because of the high availability of affiliation data in those subject areas,) than the humanities and social sciences, so an automated approach for humanities and social sciences was not feasible.

Consideration 2: Familiarity With Open Access

Open access provisions are a common requirement of funder mandates for science, technology and mathematics fields, most notably those of the National Institute of Health.¹ We postulated that due to these requirements, large

science, technology and mathematics repositories such as PubMed could be fertile sources of publicly available UNC-Chapel Hill authored content. Humanities and social sciences fields lack an overarching NIH-style mandate and a large PubMed-like repository, so we were unable to harvest humanities and social sciences content in the same way.

Consideration 3: Type of content

The Open Access Policy covers articles, which are a primary means of scholarly communication for science, technology and mathematics researchers. The policy gives us more leeway to make article content openly available in the Carolina Digital Repository. While humanities and social sciences researchers also communicate their findings via articles, they also author books, book chapters and other types of content that is not covered under the Open Access Policy. Thus, we needed to determine whether to collect non-Open Access Policy eligible content for the purposes of this pilot project.

Consideration 4: Team goals

We also needed to consider the goals of the Open Access Implementation Team when developing approaches to ingesting content into the Carolina Digital Repository. The Team wanted to focus on collecting humanities and social sciences content because these areas were not well represented in the Carolina Digital Repository and would be less likely to be available in another open access repository.

Pilot Project Structure

After identifying these goals, we realized that the differences between humanities and social sciences content necessitated different outreach approaches. To discover humanities and social sciences content, we needed to take a hands-on approach and engage with researchers and faculty one on one. Conversely, discovery of science, technology and mathematics content could be more fully automated. To address these considerations, we developed three approaches, which we divided into three pilot projects.

¹ <https://publicaccess.nih.gov/policy.htm>, accessed December 10, 2018

Approach 1: Highly Cited Authors

Since 2014, Clarivate Analytics has published an annual list of Highly Cited Researchers, who are prominent in their field and university.² For example, several of the authors employed at UNC-Chapel Hill run their own laboratories and employ researchers and graduate students. We posited that if these highly cited researchers were made aware of the Carolina Digital Repository and the Open Access Policy, they might pass along this knowledge to their employees and colleagues. Additionally, their highly cited work might drive additional traffic to the Carolina Digital Repository.

Initially, we were curious how many of these researchers were employed at UNC-Chapel Hill and whether their work was already in the Carolina Digital Repository. After examining five years worth of Highly Cited Researchers' lists, we identified 56 researchers who were affiliated with UNC-Chapel Hill.

Using Scopus, we discovered and downloaded a list of each researcher's publication history. We limited our scope to articles which Scopus listed as having over 300 citations, as this was a reasonable threshold for all 56 authors' citations and was manageable with our workload. After applying this threshold, we obtained 842 articles for further investigation.

Next, we looked at the journals that published each of the 842 articles and checked their self-archiving policies in SHERPA/RoMEO, a publisher open access policy aggregator, to determine if it was possible to deposit a copy of the article in the Carolina Digital Repository. After we identified articles which were eligible for deposit, we contacted the researcher and asked for permission to deposit a copy in the Carolina Digital Repository. To date, 286 articles from 21 researchers have been deposited. Happily, only 2 articles had been deposited in the Carolina Digital Repository prior to this project.

We were curious to see how many of the 842 articles were already available in an open access repository prior to deposit in the Carolina Digital

Repository which aligned with our goal to make content available that was not already openly available. To determine whether an article was available openly, we used Unpaywall's Simple Query Tool.³ The Simple Query Tool matched DOIs for the articles on our list to Unpaywall's vast database of openly available articles. More specifically, if the Simple Query Tool returned a DOI match, the article was available openly from another source. At the time of this analysis, the Carolina Digital Repository was not listed as a source in Unpaywall, eliminating the danger of overlap. Using this method, we determined that approximately 46% of articles on our list were not openly available prior to deposit in the Carolina Digital Repository.

Approach 2: 1Science and Author Citations

As we were considering our approach to identifying highly cited authors, our library leadership had purchased a 1foldr report from 1Science.⁴ This report identified 47,281 open access articles that were authored by UNC-Chapel Hill faculty and staff from 1980 through May 2018. The article list was compiled from open access sources including PubMed Central, publisher websites, and other open access repositories. In addition, the report filtered out articles which already appeared in our repository and checked SHERPA/RoMEO for each article's deposit restrictions. The report was a one-time only purchase, so it would not help us identify content going forward. Furthermore, the report identified articles that were already open access, rather than paywalled articles to which we could apply the Open Access Policy. The 1foldr report seemed to be best suited to filling our backfile and we determined that we needed a different approach to identify post-2018 and paywalled content.

To identify content going forward, we decided to trial a method to gather articles and their metadata using search alerts and RSS feeds. We adapted this process from work developed at Montana State University and the College of Wooster, among others.⁵

² See <https://hcr.clarivate.com/>

³ See <https://unpaywall.org/products/simple-query-tool>

⁴ 1Science has since been acquired by Elsevier.

⁵ For more information, see: Sterman, L.B. and Clark, J.A. (2017). Citations as Data: Harvesting the Scholarly Records of Your University to Enrich Institutional Knowledge and Support Research. *College and Research Libraries*, 78(7), 952-963, DOI: 10.5860/crl.78.7.952 and Flynn, S.X.; Oyler, C.; and Miles, M. (2013). Using XSLT and

Next, we set up institutional affiliation search alerts in Scopus and Web of Science and keyword search alerts in Google Scholar for UNC-Chapel Hill. Since Scopus and Web of Science normalize affiliation data, but Google Scholar does not, data originating in Google Scholar needed to be manually checked to verify that the flagged article was actually written by a member of the UNC-Chapel Hill community. Metadata for 5,724 articles was saved to Zotero, examined for duplicates, and uploaded to a shared Google Sheet. We then ran a script developed by the College of Wooster to query SHERPA/RoMEO for the journal's deposit policies.⁶ We plan to manually query the university directory for the UNC-Chapel Hill affiliated author's contact information and ask for their permission to deposit.

This approach identified both content that was openly available and behind paywalls, which met the goals of the Open Access Implementation Team. However, due to the reliance on affiliation data for article identification, the results were highly focused on science, technology and mathematics. Additionally, we found that this approach is very labor intensive; the time needed to import, de-duplicate and normalize the metadata records is considerable for our small staff. It is likely that extensive staff time will also be needed to manage the author permissions and upload process.

Approach 3: CV Review

Lastly, from anecdotal reports, we suspected that a barrier to faculty deposits, particularly for pre-2016 publications, is the level of effort and time that it would take to locate legal versions of articles and research publisher archiving policies. In addition, the time it takes to complete the deposit workflow for individual articles could also be a deterrent. As a pilot initiative, our goals were to 1) increase the number of articles from arts and humanities faculty, 2) establish an understanding of the time commitment to complete this work, and 3) determine if this was a sustainable service that the Open Access Implementation Team could provide going forward.

In coordination with the liaison librarians, we had the opportunity to visit six academic departments

between January and September 2018. First, we contacted the department chairs and asked for five to ten minutes at their upcoming departmental meeting to share information on new open access initiatives on campus. Once we agreed on a date with the department chair, we invited their liaison librarian to join us at the faculty meeting. For each department, we provided overview information about the Carolina Digital Repository and a tailored message about open access publishing in their discipline. We also referred back to the campus Open Access Policy as a way to show institutional support. After we had established the context, we asked faculty members to participate in the CV Review project by emailing us a current copy of their CV, either as a PDF or a web link. The benefit for faculty would be to see which of their publications were eligible for deposit into the Carolina Digital Repository, and then with their permission we would deposit the articles for them. We also explained that the benefit to us would be to increase the research available in the Carolina Digital Repository and also allow us to test the workflows and level-of-effort needed to sustain this service. In addition, we left time for questions from the groups in order to engage with the faculty and hear their concerns. Following the meeting, we emailed the chair with a reminder for faculty to share their CVs with us, links to open access information on the Libraries' website, and our contact information.

In total, we received only four CVs from faculty, and two of those were from departments other than the ones we visited in 2018. With help from a graduate field experience student, we reviewed the CVs and found that the majority of articles that were eligible for deposit into the Carolina Digital Repository were already accessible through PubMed or as gold open access articles. In addition, the average time to process a CV of an associate professor was ten hours. Although when reviewing only the subset of articles published since 2016, the time needed was two hours on average. We received no CVs from arts or humanities faculty.

The results of the small CV Review pilot indicated that this was a very labor intensive process, but that

Google Scripts to Streamline Populating an Institutional Repository. *Code4Lib Journal*, 19. Retrieved from: <https://journal.code4lib.org/articles/7825>

⁶ See Appendix A of Flynn et al. for the source code.

had given us a useful outreach tool for faculty. However, as only a few faculty took advantage of this opportunity and the resources we discovered through the process were within the Carolina Digital Repository's scope, we will continue to offer this service by request.

Lessons Learned

Over the course of the year that we implemented the Content Liberation project, we learned a great deal about the open access landscape at UNC-Chapel Hill and gained a better understanding of the needs and concerns from faculty across campus. Notably, the two main themes were:

- Faculty were unlikely to complete the manual deposit process.
- Across disciplines, tenure level, age, and gender, faculty were confused about open access, unaware of the Open Access Policy, and unsure what information they needed.

However, these challenges presented a number of opportunities for us to engage in continued outreach with faculty. For example, although faculty did not have time or desire to deposit their articles in the Carolina Digital Repository, they were almost always willing to give permission for us to do so on their behalf. In some instances, we were able to deposit the articles that we found through the Content Liberation project and additional articles provided by the faculty member that were within the Carolina Digital Repository's scope, but were missed through our processes. In addition, we were able to continue working with faculty by maintaining conversations around open access resources for classroom instruction, providing direct guidance about the open access policy, and leading several targeted workshops for faculty on open access information. We also saw the value of frequent communications around open access services and the Carolina Digital Repository specifically, in order to help remind faculty and keep them engaged.

Future Directions

Our top priority is to launch our new institutional repository platform, which is based on Samvera's Hyrax solution. This project will split our current repository into two systems: an institutional

repository and a special collections repository, which allows us the freedom to customize each system for their specific use case. The new institutional repository system is optimized for scholarly content and contains out of the box features that we feel will be attractive to depositors, including search engine optimization, download statistics, single use links, collections, and proxy deposit. Additionally, we plan to mint DOIs for all repository content. Hopefully, these new features and services will encourage users to deposit their work and will increase discovery of the repository's content.

Additionally, we would like to expand our data sources for UNC-Chapel Hill authored content. Our current sources are heavily weighted towards science, technology, and mathematics aggregators, and we'd like to include more humanities and social sciences content. We also plan to explore methods to make the current content identification, investigation, and repository ingest process integrate further with our new repository platform.

In tandem with the launch of the new repository platform, we are planning a multi-pronged outreach campaign for faculty to promote the Libraries' open access services. To date, we have completed the initial market research and begun developing the open access brand identity. We will host a series of kick-off events and consultations to drive interest in open access activities and provide information on the practical components.