

Content Liberation Project Summary

Rebekah Kati
Institutional Repository Librarian
University of North Carolina at Chapel Hill

Introduction and Project Goals

The UNC Libraries Open Access Implementation team has been charged with increasing the amount of content in the Carolina Digital Repository (CDR) and raising faculty awareness of the UNC Open Access Policy and author rights issues. The team has determined that providing access to legal versions of paywalled content will best address the charge and will demonstrate the value of the CDR to the faculty. Therefore, the team has developed an overarching project, Content Liberation, to extract content from behind paywalls.

The general goals for the Content Liberation project are:

- Fill the repository with high-quality content
- Raise faculty awareness of the Open Access policy
- Identify areas of repository and service growth

Project Approaches

Faculty research at the University of North Carolina at Chapel Hill is strong in both humanities and science content. The Open Access Implementation Team identified content from both subject groupings, but soon realized that different approaches were needed for extraction. Therefore, the overall project has been divided into three parts: Author Citation Search, CV Review and Highly Cited Authors. Each part of the project addresses the acquisition of a different content subject grouping. Over the course of the project, the Open Access Implementation team will work with the Repository Applications Team to identify programmatic methods to automate repetitive tasks.

Author Citation Search

To identify articles written by UNC faculty in STM fields, the Open Access Implementation Team will conduct affiliation searches in the library databases. This approach easily identifies researchers from UNC and aggregates many citations quickly.

The searches will be conducted primarily in Scopus and Web of Knowledge. Additional databases such as Project Muse and Academic Search Complete will be used as needed. Since UNC Libraries has already purchased access to these databases, there is no extra cost associated with the search. The search criteria are UNC authored publications from 2016 to the present. 2016 represents the year that the Open Access Policy went into effect.

Once appropriate citations have been located, they will be imported into Zotero for normalization. The citations will then be exported into a Google Sheet. CDR staff will then fill in empty metadata fields,

locate the UNC author (if the article was written by multiple authors), identify the UNC author's contact information and run a Sherpa/ROMEO API search on the journal's ISSN to reveal the journal's deposit policy category. After the article's deposit policy has been located, the team will contact the author and walk them through the deposit process or deposit the article on their behalf.

Advantages

- Large number of citations can be located quickly
- STM focus aligns with greatest faculty exposure to open access journals and terms, providing an existing base of knowledge that the team can build on
- Metadata can be downloaded and modified quickly for load into the CDR

Challenges

- Citations are primarily from STM journals due to the subject area focus of Scopus and Web of Knowledge, as well as publisher conventions regarding affiliations
- In the case of multiple-authored articles, staff time will be spent identifying the UNC author
- Staff will need to spend time locating the author's contact information
- Staff will need to spend time verifying the terms of deposit against Sherpa/ROMEO, as the API returns general deposit categories rather than nuanced terms of use.

Area of growth

- Currently, Scopus and Web of Knowledge impose a download limit when searching from the web interface. This is not anticipated to be a critical issue since the project is focused on post-2016 content. The official APIs could be used to obtain additional content, should the team wish to explore a more programmatic approach to citation imports, or obtain pre-2016 content.
- The searches could be expanded to additional databases if the team wished to broaden the scope, or target a subject area.

CV Review

The team will partner with the Library's Research and Instructional Services staff and university departments to collect and review faculty CVs. Research and Instructional Services librarians will make appropriate introductions to receptive departments on the team's behalf. Open Access Implementation team members will attend a faculty meeting to give a short introduction to the service and then invite faculty to submit their CVs for review.

After the CVs are received, team members will look up each citation, extract and normalize the metadata via Zotero, import the metadata into a Google Sheet and run a Sherpa/ROMEO API search on the journal's ISSN to reveal the journal's deposit policy category. After the article's deposit policy has

been located, the team will contact the author and walk them through the deposit process or deposit the article on their behalf.

Advantages

- Teach principles of open access and author rights directly to faculty
- Obtain lists of faculty publications from humanities and social sciences disciplines
- Develop ongoing relationships with university departments

Challenges

- Faculty also publish books, book chapters and conference proceedings that are not covered by the open access policy. Since these materials are not covered by Sherpa/ROMEO, locating the publisher's deposit policy could be difficult
- Initial barriers to departmental contact and faculty compliance are high
- Staff will need to spend time verifying the terms of deposit against Sherpa/ROMEO, as the API returns general deposit categories rather than nuanced terms of use.

Areas of growth

- Expand program to additional subject areas and/or populations, such as graduate students and adjuncts
- Investigate integration opportunities with departments. For example, new staff members could be approached and encouraged to submit their CVs.
- The team could track faculty publications and contact faculty when they publish a new article

Highly Cited Authors

Clarivate Analytics (the parent company of Web of Knowledge) publishes a list of [Highly Cited Researchers](#) once per year. This list shows the top cited researchers in science and social science over a ten-year span and includes researchers from UNC. Previous years' lists are available.

To drive traffic to the repository and make high-quality, high-use scholarship available, the team will extract UNC-affiliated authors from the available lists. Once a combined list is generated, the team will run author publication searches in Scopus. Scopus was selected because it is a comprehensive indexer for the researcher's subject areas and returns citation counts. The publication list will then be imported into a Google sheet and a Sherpa/ROMEO API search will be run on the journal's ISSN to reveal the journal's deposit policy category. After the article's deposit policy has been located, the team will contact the author and walk them through the deposit process or deposit the article on their behalf.

Advantages

- Easily obtain a list of high-quality scholarship from prominent UNC researchers
- Highly-cited content is likely to bring additional traffic to the CDR and increase awareness

Challenges

- Author publication history search in Scopus returns all publications over the author's career, including those authored while the author was employed outside of UNC
- Highly Cited Authors lists are weighted heavily towards STM content
- Author publications are generally pre-2016 and outside the scope of the open access policy
- Staff will need to spend time verifying the terms of deposit against Sherpa/ROMEO, as the API returns general deposit categories rather than nuanced terms of use.

Areas of growth

- Explore other sources of highly cited content. Do other analytics companies publish lists?
- Review current year's list when it is released

Outcomes and Timeline

The outcomes of the overall project are:

- Increased deposit of high-quality STM and HSS content in the CDR
- The Open Access Implementation team will attend select faculty meetings
- The Open Access Implementation team will create a plan for extending services in the future

Each of the three parts of the project will be piloted from January-March 2018. The first batch of content will be loaded into the CDR by end of Spring Semester 2018. The team will attend faculty meetings through the end of Spring Semester 2018. The plan for future growth will be finalized by the end of August 2018.