

Content Liberation Update, August 2020

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

Introduction

In 2018, the Open Access Implementation Team (the Team) conducted three Content Liberation pilot projects to ingest UNC-authored content into the Carolina Digital Repository (CDR). A summary of each project is available in [Content Liberation Overview](#).

This report summarizes our current progress on these projects and suggests recommendations for progress in the next academic year. A follow-up report will be available in August 2021.

- [August 2018 Content Liberation Update](#)
- [August 2019 Content Liberation Update](#)

Author Citations (Green OA)

Work on the Author Citations project has restarted, and the project has been renamed “Green OA.” In November 2019, the Software Development and Repository Services Departments met to discuss potential collaboration and automation strategies. The conversations centered around potential data sources which could be utilized to import UNC-authored content into the CDR automatically. We identified four potential data sources and assessed them against a sample set of 100 articles: Web of Science, CrossRef, Scopus and ORCID. Assessments have been completed for CrossRef and Scopus. Web of Science did not reply to our repeated queries for an API key, so their content has been excluded. Since ORCID’s API relies on user input for content, it was not as complete as the other sources. So far, it seems that the Scopus API contains the most complete metadata in a format that could be imported into the repository.

Additionally, we assessed the large collection of saved citations that the Institutional Repository Librarian has been collecting since late 2017. Repository Services hired a Research Assistant to help with data entry and cleanup for the Author Citations spreadsheets. Once the spreadsheet is complete, we will start contacting authors and devising a strategy for upload.

1foldr

As reported in the 2019 update, the Team began the process to upload 1foldr content written between 2016 to 2018. During our assessment, we found several issues that required remediation before content could be loaded into the CDR, specifically:

- Author names appeared in several places and were not consistent
- The author departmental affiliation metadata, when present, did not conform to our internal metadata schema
- Embargo information was not provided by 1foldr

- A bug in the 1foldr download script overwrote article file names, which caused PDFs to be matched to incorrect metadata records. Our developers needed to rewrite the script to fix this issue and we needed to re-download the batch

After multiple iterations, we loaded over 5000 articles from the 1foldr report into the CDR in April 2020. This first batch represents articles which are covered under UNC's Open Access Policy. This batch helped us to achieve our goal of refocusing the content of the repository from student papers to faculty scholarship, as our article resource type counts surpassed our student paper counts for the first time.

After a copyright consultation, we identified a second set of approximately 22,000 articles from 1980-2015, which were available in PubMed. Since the articles were already available publicly, it was determined that they would be appropriate to deposit in CDR. Our extensive work on the previous batch has enabled us to streamline loading of the current batch and we hope to load the rest of the articles before the end of Fall semester 2020. After the current batch is loaded, we will assess the remaining articles to determine which versions we are legally allowed to upload.

CV Review

In March, the COVID-19 pandemic necessitated that many of the Libraries' staff start to work from home. In response to that need, the Institutional Repository Librarian adapted the CV Review Content Liberation project so that it could be completed by library staff and students who might be unfamiliar with scholarly communication and open access. As of August 1, 2020, the CV Review Work from Home team has reviewed 212 CVs containing 9175 articles. 2864 of those articles are eligible or deposit in the CDR. This project will run throughout Fall 2020 as at-home work is needed during the COVID-19 pandemic. Many of the workers on the project have chosen to focus on humanities and social sciences departments and which will help us diversify the subject areas covered in the CDR.

Once it became clear that UNC would be a key research institution contributing to COVID-19 research efforts, the Institutional Repository Librarian created a CV review sub-project to ensure that outputs of coronavirus research were available in the CDR. Using a list of researchers provided by the School of Medicine, the Institutional Repository Librarian and a research assistant reviewed each researcher's CV, contacted them and deposited the eligible articles. This project drew the attention of UNC's Communications department and the project was profiled in *The Well* along with the Library's open access efforts. This approach demonstrates how Content Liberation projects can be utilized to address issues of national concern and to ensure that relevant research is appropriately targeted, deposited and preserved.

Highly Cited Authors

In December, we once again looked at Clarivate's list of Highly Cited Authors and identified 11 new authors who had not been previously contacted. We reviewed 94 articles and found 17 which were eligible for deposit. The Team defined "highly cited article" as an article which received over 300 citations in Scopus. We sent a list of eligible articles to each author and asked them to contact us if they did not want the articles to be deposited into the CDR. This was a change from previous years in which we took an "opt-in" approach.

The Team performed an assessment to determine whether any of the 94 articles were already available openly prior to deposit in the CDR. Using Unpaywall's Simple Query Tool, we found that:

- 34 articles were not available openly in a repository or on the publisher's website (36.2% liberated)
- 60 were already available openly in a repository or on the publisher's website (63.8%)

Additionally, we also determined whether the 94 articles were available on an academic social network or author's institutionally affiliated website:

- 26 articles were not available on an academic social network or an author's institutionally affiliated website (27.7%)
- 68 were available on academic social networks or an author's institutionally affiliated website (72.3%)

Although much of the scholarship is open elsewhere, we feel that it is important enough to preserve in the CDR. Additionally, the project will raise awareness of the CDR to highly influential faculty, which could lead to increased deposits in the future. Three of the researchers from the 2019 group asked for their full CV to be reviewed.