Carolina Digital Repository Update 2020

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

Contents

Introduction

The Carolina Digital Repository (CDR) is a Samvera Hyrax-based institutional repository, which launched on June 5, 2019. Prior to June 5, 2019, the CDR platform was built on custom Fedora and housed both institutional repository and born-digital special collections materials. The former CDR platform now contains only born-digital special collections materials, as well as one highly customized institutional repository collection (Research Labs of Archaeology). Additionally, the UNC-Chapel Hill Libraries administers a CONTENTdm platform for digitized special collections content.

This report summarizes activity in the CDR from January 1, 2020 through November 30, 2020. It describes key initiatives, growth and usage of the repository as well as a description of future initiatives. We would like to thank Joy M. Perrin at Texas Tech University for the inspiration and format of this report.

Key Initiatives: 1Science content

In May 2018, UNC Libraries purchased a report from 1Science, which listed 47,281 articles which were available openly and might be eligible to deposit in the CDR. Due to the size of the dataset and potential rights issues, we did not want to load all 47,281 articles into the CDR at once and elected to separate them into batches:

- Batch 1: Articles published between 2016-2018 and covered by [UNC-Chapel Hill's Open Access Policy](#)
- Batch 2: Articles published between 1980-2015 and openly available in PubMed
- Batch 3: Articles published between 1980-2015 and not available in PubMed

When we prepared to load Batch 1, we noticed issues with the vendor-provided metadata, such as the lack of author order and omission of departmental information. We decided to use several metadata fields from Scopus in place of 1Science's metadata to correct these errors and omissions. This required substantial development work, metadata reconciliation and several rounds of quality assurance testing. Anna Goslen and Rebekah Kati [presented on this project](#) at the 2020 Triangle Research Libraries Network Annual Meeting.

We loaded Batch 1 in April 2020, which consisted of 5,638 articles. Using the lessons that we learned from Batch 1, we loaded 22,857 articles at the end of October 2020. We plan to load the third batch of approximately 5,000 articles in 2021. The remaining articles will need to be individually vetted for deposit permissions, which will slow the pace of the work considerably.

Key Initiatives: COVID-19 Authors and CV Review projects

In March 2020, UNC-Chapel Hill pivoted to online learning in response to the COVID-19 pandemic. In response to this pivot, the University Libraries needed to provide work for Library staff and students who were working remotely. We developed the CDR CV review project in response to this need. In this project, Library staff and students chose a department and located CVs of faculty members within that department. They evaluated all the articles listed on the CV for eligibility for deposit in the CDR. From March through November 2020, the staff members and students have evaluated 251 CVs and reviewed 4,436 articles. 1,311 of the articles reviewed are eligible for deposit in the CDR. Throughout the next year, we will load the eligible articles into the CDR.

At the start of the pandemic, it became clear that UNC-Chapel Hill was a leading institution for coronavirus research. In support of the researchers and the UNC-Chapel Hill Library's Sustainable Scholarship initiative, we developed a plan to review CVs of faculty involved in coronavirus research and deposit eligible articles in the CDR. Using UNC-Chapel Hill news sources, we identified 85 researchers who are currently working on coronavirus-related research. They have authored 10,825 articles, 4,003 of which are eligible for deposit. As of this writing, we have contacted all authors and have loaded 72 authors' eligible articles into the CDR. We will continue to load the rest in the coming year. Additionally, we've set up search alerts in PubMed for COVID-19 research and are monitoring news sources for

additional researchers to approach. All research specifically related to coronaviruses is in a dedicated collection: UNC-Chapel Hill Coronavirus Research. This project was profiled in the UNC-Chapel Hill newsletter *The Well* in April 2020.

Key Initiatives: Bulk Import and Export

The CDR's platform, Samvera Hyrax, contains many useful features out-of-the-box, but it does not contain a feature to upload content to a repository in bulk. Consequently, we had to upload content to the CDR individually via the web form or request developer time to customize an upload script. This year, we investigated two bulk upload solutions for Hyrax, Ziza and Bulkrax. We chose Bulkrax, as it seemed to be farther along in development and was more popular within the Samvera community. In May, we launched Bulkrax in CDR and it has vastly improved our ability to upload content to the system quickly and efficiently.

Key Initiatives: Optimize CDR metadata for Google Scholar

Although Samvera Hyrax is optimized for Google out-of-the-box, additional work was needed to optimize the CDR for Google Scholar. We added additional metadata to each work's HTML header which enabled Google Scholar to index volume, issue and page numbers. We also implemented author ordering, reindexed articles to reflect author order and performed a QA process to ensure that the author order was working correctly. Monica Westin at Google was an instrumental resource in this effort.

These efforts paid off. In 2019, CDR received 1,578 referrals through Google Scholar. In 2020, CDR received 47,080 referrals through Google Scholar. We anticipate additional referrals in 2021 as we keep adding more articles to the CDR.

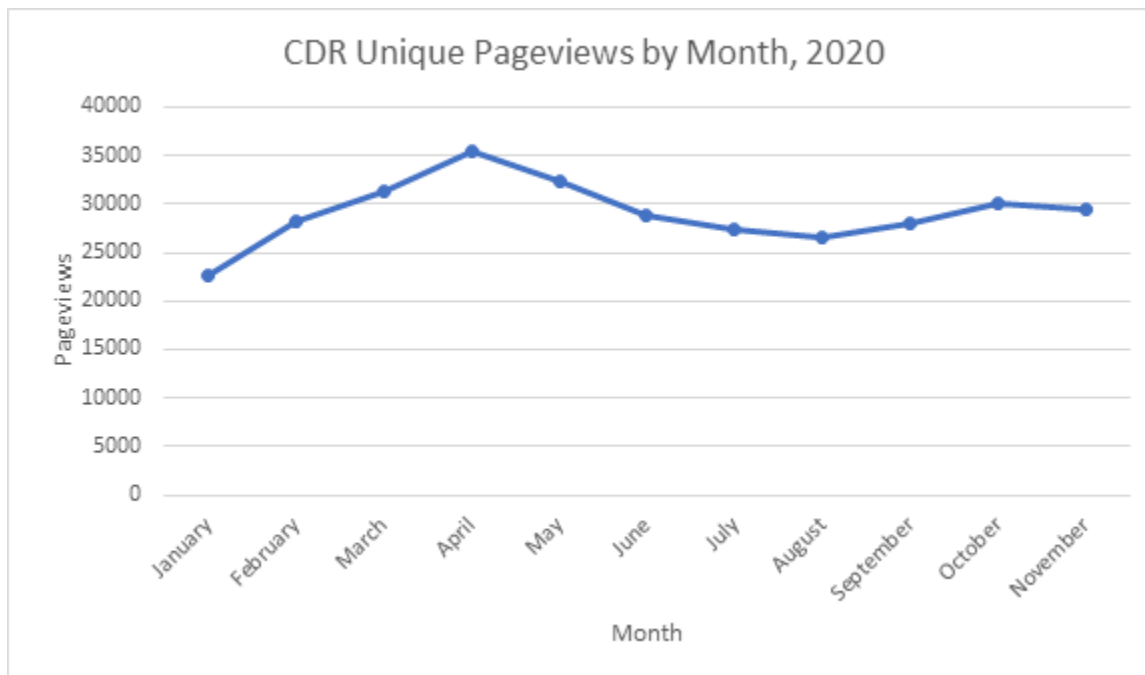Other features and initiatives completed this year include:

- Implement author order on work show pages and in metadata
- Development and implementation of policy to remediate CDR content for accessibility by request
- Bug fixes
- Platform upgrades
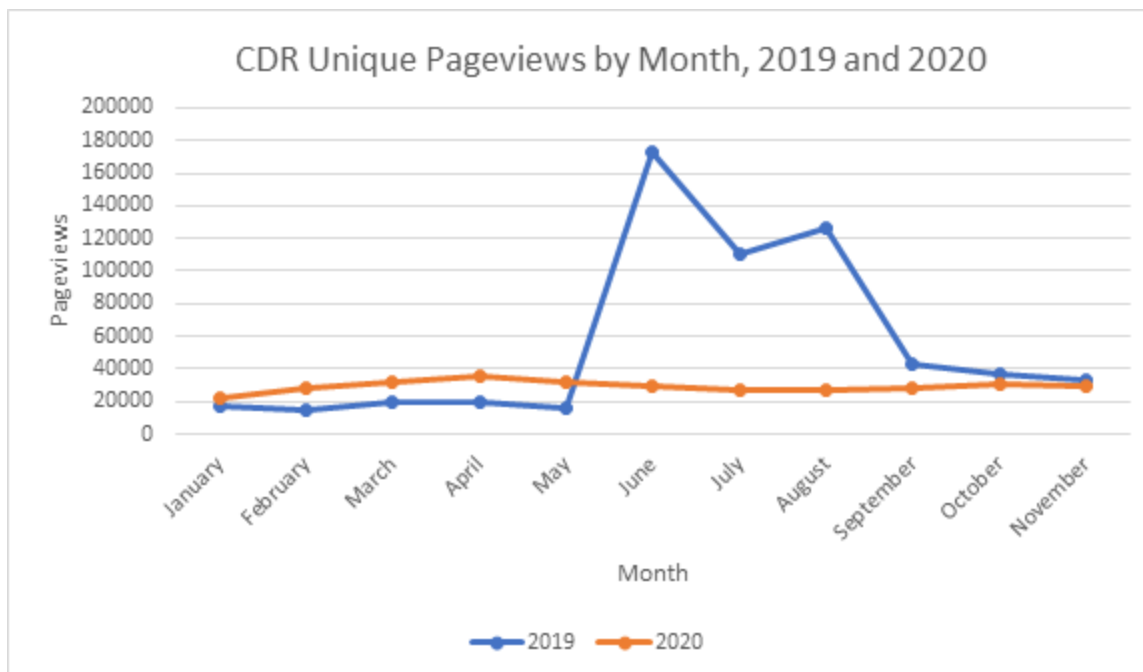- Several metadata remediation projects

Growth

This year, we added 32,424 works to the CDR. Most of these works were articles which were added via the 1Science report, which is explored in-depth above. The impact of the 1Science batches should not be overlooked. Prior to the launch of the new CDR in 2019, the system contained 6,834 articles. Due to the approximately 28,000 articles added from the first two 1Science batches, we have effectively quadrupled the number of scholarly articles available in the CDR through this single initiative. Although the 1Science content was available openly elsewhere, it helped to change the focus of the CDR from a repository focused on student work to one focused on faculty scholarship.

Usage

The CDR recorded 320,428 unique pageviews between January 1, 2020 and November 30, 2020. Our busiest day was Monday, March 2, 2020 during which we received 3,406 pageviews.
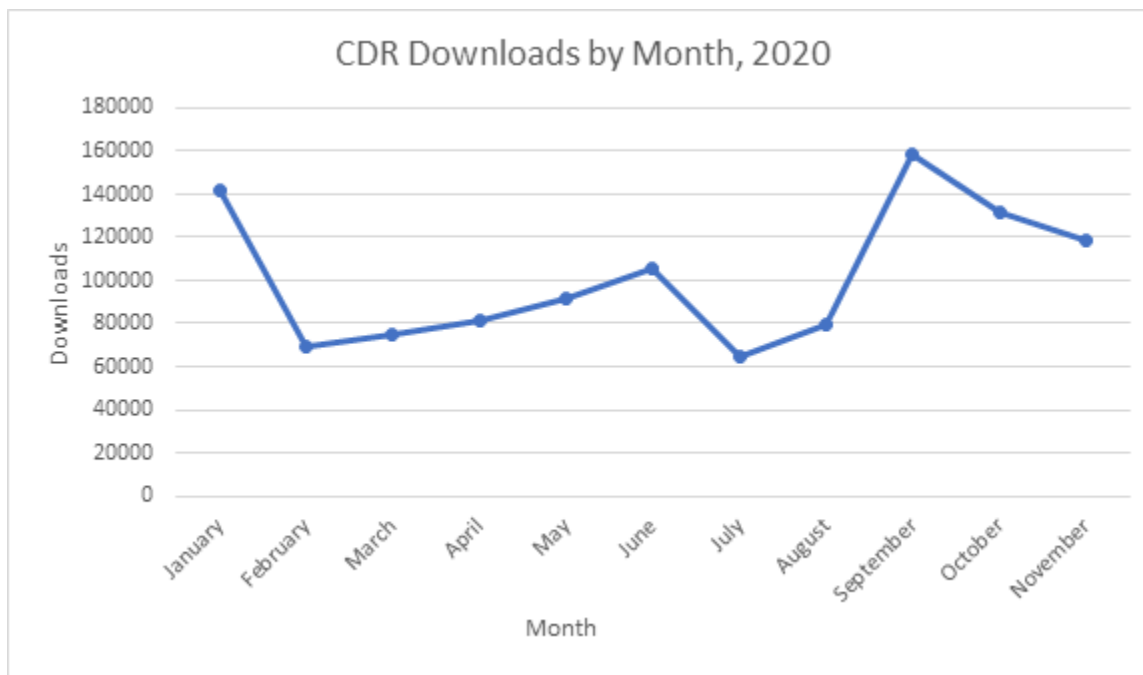
CDR Unique Pageviews by Month, 2019 and 2020

The large spike in June-August 2019 was due to the popularity of the SILS masters paper "Predicting the Difficulty of Trivia Questions Using Text Features", which received widespread media attention due to the author's appearance on *Jeopardy* in June 2019. After the media attention waned, the CDR's usage stats fell back to normal levels. CDR received more pageviews in January through May 2020 than during the same period in 2019, which can be attributed to increased search engine optimization from our new platform.

Most Viewed Works, 2020

Work Title	Resource Type	Number of Unique Pageviews
Effects of heterogeneous and homogeneous grouping on student learning	Masters Thesis	1,632
Filled-in vs. Outline Icons: The Impact of Icon Style on Usability	Masters Paper (SILS)	1,572
Social Media Use and Its Impact on Body Image: The Effects of Body Comparison Tendency, Motivation for Social Media Use, and Social Media Platform on Body Esteem in Young Women	Masters Thesis	1,488
On-Campus Parking: Problems and Solutions	Masters Paper (City and Regional Planning)	1,291
Struggling to Breathe: COVID-19, Protest, and the LIS Response	Article	1,272
Walking on Eggshells: Experiences of Students of Color within Library and Information Science Master's Programs	Masters Paper (SILS)	1,069

Tom's Comments: Advice about Graduate School, Finding a Job, Reaching Tenure in Political Science and Other Social Sciences, and All of the Steps in Between	Book	982
Constructive Journalism: The Effects of Positive Emotions and Solution Information in News Stories	Dissertation	976
Shadows and stadium lights: an inside look at the 2009 UNC women's soccer team	Masters Thesis	907

The CDR received 1,118,052 downloads during the same time period. This represents an increase of 479,588 downloads from 2019. Note that downloads for part of March and all of April and May 2019 were not recorded due to technical issues, so the actual difference between 2019 and 2020 may vary.

Most Downloaded Works, 2020

Work Title	Resource Type	Number of Downloads
Social Media Use and Its Impact on Body Image: The Effects of Body Comparison Tendency, Motivation for Social Media Use, and Social Media Platform on Body Esteem in Young Women	Masters Thesis	7,317

Born Again Digital: Exploring Evangelical Video Game Worlds	Dissertation	6,254
A Detailed Journey into the Punk Subculture: Punk Outreach in Public Libraries	Masters Paper (SILS)	4,737
Global to Channel Scale Water Level Forecasting and Analysis	Video	3,921
The nature and determinants of customer expectations of service	Article	3,819
The Kongo Rule: The Palo Monte Mayombe Wisdom Society	Dissertation	3,804
Improving the Accuracy of a Real-Time ADCIRC Storm Surge Downscaling Model	Video	3,703
Modeling the Storm Surge and Compound Flooding from Hurricane Florence Using ADCIRC	Video	3,635
ADCIRC Version 54	Video	3,604
ADCIRC v55 Modeling the Earth, Mesh Resolution Effects and Removing Time Step Constraints	Video	3,588

The CDR has more downloads than pageviews, which shows that users encounter CDR content through venues other than the CDR's interface. The top referrers to CDR content include:

- Google Scholar
- Facebook
- UNC LibGuides
- Baidu
- UNC Library Catalog

Collection Highlights

We continue to support student paper deposit, as well as self-deposit of faculty work. Our notable new partnerships in 2020 included:

ADCIRC Users Group Meetings

ADCIRC is a software program developed by the UNC Institute for Marine Sciences. Due to COVID-19, they deposited their 2020 User's Group meeting recordings in the CDR, which were downloaded thousands of times.

Moore Undergraduate Research Apprentice Program (MURAP) Projects

Students participating in the MURAP fellowship program deposited their final projects in the CDR.

On the Books: Jim Crow and Algorithms of Resistance

In support of an Andrew W. Mellon Foundation Grant, the CDR is storing the XML files and white paper for UNC Libraries' On the Books project.

Research Laboratories of Archaeology Publications

The RLA Publications collection was moved from the Digital Collections Repository to the CDR at the project leader's request. In addition, we added new publications and enhanced the metadata for existing publications.

Series on Germanic Languages and Literatures

In support of a National Endowment for the Humanities Grant, the CDR provided access to 120 books which address scholarship in the Germanic Languages. The books were originally published by UNC Press and the UNC Department of Germanic and Slavic Languages and Literatures.

Southern Sources: A Symposium Celebrating Seventy-Five Years of the Southern Historical Collection

The Southern Sources collection contains research papers written in celebration of the Southern Historical Collection. They were originally stored in UNC Libraries' CONTENTdm repository, but were moved to the CDR to accommodate more

Faculty Work

We've made great progress this year preserving faculty work due to our 1Science, COVID authors and CV review projects detailed above. As in 2019, faculty are not self-depositing into the CDR at a sustainable scale. Therefore, it is important that CDR staff continue to identify and deposit faculty work. We have continued the Highly Cited Authors project, which aims to deposit high-impact research from UNC authors into the CDR. The GreenOA project aims to identify recent, paywalled articles authored by UNC researchers and deposit them into the CDR. In support of the GreenOA project, we have hired a Graduate Assistant who is cleaning up metadata, locating faculty contact information and obtaining article PDFs. Both projects utilize mediated deposit rather than self-deposit, which we hope will increase uptake.

Discussion and Future Steps

Over the past year, we focused our efforts on feature development and enhancements for the CDR, particularly to optimize the repository for Google Scholar and other indexers. The analytics show that these efforts have been successful, as CDR is receiving much more referral traffic from Google Scholar than it did in 2019. Our pageview traffic has levelled off after an anomalous 2019, and we are seeing consistently higher pageviews and downloads than we did on the previous CDR platform. We also implemented a batch upload tool, which has enabled CDR staff to speed up the article upload process.

Progress over the next year will be slow. We expect that next year's upload numbers will be much lower, due to staff turnover and the need to do individual rights vetting for GreenOA and the remaining 1Science articles. Due to the University's COVID-19 related hiring freezes, continued vacancies will affect our work. Despite these challenges, we anticipate that we will continue to focus our efforts on faculty scholarship by loading articles from the CV review project and the GreenOA project. Furthermore, in support of the Library's diversity, equity and inclusion initiatives, we will undertake an audit of the CDR's content initiatives to uncover and document inequities.