Analyzing Institutional Open Access Publishing Activity

William Joseph Thomas

Abstract

Scholarly communication librarians routinely promote Open Access (OA) publishing, encouraging their institutions' authors to increase their impact for the benefit of a greater number of readers. Researchers at East Carolina University (ECU) collected six years of ECU authors' publication data to analyze how much per year was free to read, and how much more could be made free for end-users to read. To do that, bibliographic information about the publications was merged with OA information provided by Unpaywall and Sherpa/RoMEO. Results were collated by author to target potential OA champions and for outreach to encourage green OA deposit when permitted. ECU's top journals, publishers, and funders were identified and our institution's OA rates were compared to those from multiple peer institutions. The authors also identified next steps, including adding data from 2021 and 2022 and analyzing that data for any potential effect of new OA memberships and publishing agreements, as well as continued outreach to authors.

Keywords: citation analysis, academic libraries, scholarly communication, Open Access publishing

East Carolina University (ECU) is a regional public university with about 28,000 students and 2,000 faculty members. The university received approximately $75,000,000 in funded research for fiscal year 2021. There are thirteen PhD and six professional doctoral programs. Research support is provided by Academic Library Services and the Laupus Health Sciences Library, which together maintain access to roughly 450 databases, 120,000 e-journals, 1,100,00 e-books, and about one million print titles.

This research was conducted by the assistant director for collections and scholarly communication, Joseph Thomas, working with collection development librarian Daniel Shouse. Together, they analyzed six years of publications by ECU authors to identify articles that are or could be made free to read and examined trends in the Open Access (OA) status of ECU publications. Specifically, the purposes of the study were: 1. discover how much of ECU's publishing output is free to read (gratis Open Access); 2. determine the quantity that could be made gratis OA and identify which articles they are; 3. look for evidence of growth of Open Access in any discernible trends through the six years studied; 4. identify which faculty members have the most OA publications; and 5. identify ECU's top journals and publishers during those six years.

To answer these research questions, Thomas and Shouse searched the Scopus database for all articles by authors affiliated with East Carolina University that were published from calendar years 2015 through 2020, and limited to articles with a status of "final." We exported the citation information, the bibliographical information, and the funding details for all articles. These data were collated to a master spreadsheet with a single line for each article.

Funding information was simplified by combining the text of all relevant funding fields into a single cell and then adding a column to record a single new, simplified value; for instance, funding from the National Institutes of Health or any of its institutes was recorded as "NIH." It was common for the funding fields to contain information that was not related to a funding agency, such as thanks for people who provided equipment; this information was ignored for the present study. Where there were no statements or non-funding statements, a value of "None" was recorded. Although some funding fields contained a single entity, other funding fields contained acknowledgements for multiple funders. For articles with multiple funders, the largest federal funders were preferred, and then all other funders were looked up in one or both of the registries of funder policies, Sherpa Juliet and/or Roarmap to determine whether there was an OA requirement for that funder. If any funder had a public access deposit requirement, that funder was selected for the simplified funding value. For articles that were products of funded research, but none of the funders required OA, the value selected was "None with Public Access Requirement."

Thomas and Shouse then took lists of Digital Object Identifiers (DOIs) for all articles that had them, uploaded them to Unpaywall's Simple Query Tool, and added the resulting report about the articles' Open Access status to the primary spreadsheet. There were a few mismatches between the OA status noted in Scopus and what was recorded in Unpaywall. Most of these mismatches were bronze OA articles that had reverted to subscription access. For all articles with a closed status, Shouse looked up the journal title in Sherpa Romeo for options to make the accepted manuscript or the final version available as OA. The researchers excluded the submitted manuscript options from this part of the study under the belief that the accepted manuscripts have added value because of the peer review process and because several campus authors have told us that they do not keep the submitted manuscripts as separate and distinct versions of the article.

The result set consisted of 4,965 articles by more than 2,800 ECU authors. These articles were published in 2,426 journals by 261 publishers. To get at the answers to our first three research questions, we looked at the Open Access status as a percentage of the total number of articles per year in the data set. We did not know what to expect, but found that ECU's OA publication rate ranged from 45 to 49 percent during the six years studied. There was not a clear upward trajectory, although the height of our OA

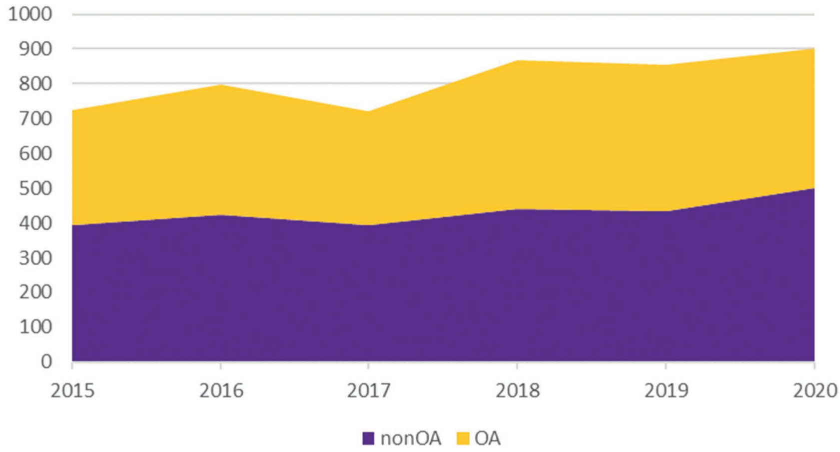

Figure 1. Free Access to ECU Articles

activity was in 2018 and 2019 before a dip occurred in 2020. We do not know the reason for the decline, whether it had to do with funding, transitions in who was publishing that final year of the data set, or a lag in public access copies becoming available (since most federal agencies allow for a twelve-month embargo). See Figure 1 for an overall breakdown of OA versus subscription access for each year studied.

To determine the maximum amount of East Carolina University's articles that could be made free to read, the researchers reviewed the Sherpa Romeo options listed for all articles that were subscription access only. Results were truly exciting: on average, only about 3 percent of ECU's article output could not be made OA. Shouse manually searched a total of 1,747 journal titles in Romeo; one hundred thirty-five journals were not listed in the database–and thus did not include OA options–and ten of the searched titles indicated that OA options were unknown. The total number of articles with this unknown or none status ranged from sixteen to thirty-four per year, or 2 to 4 percent of the total articles per year. What that means to us is that up to 97 percent of ECU's articles could be made free to read if we had the support of authors and had the capacity among library personnel

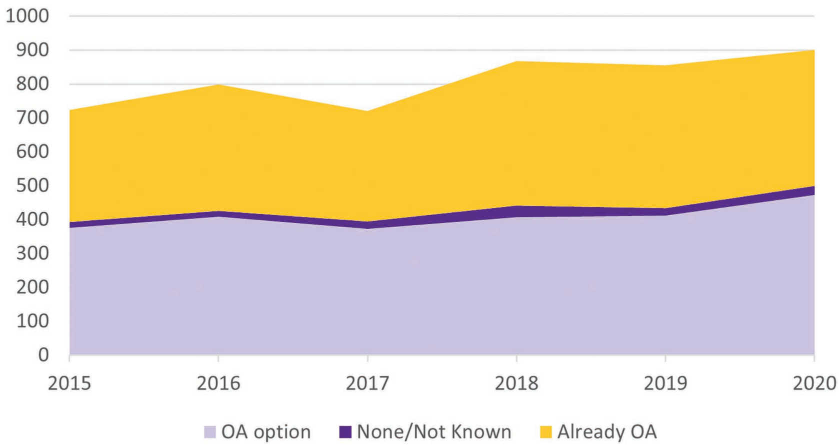

Figure 2. Additional Read Options for ECU Articles

to do the work. See Figure 2 for the maximum potential effect of OA options on ECU's subscription-access articles.

In order to celebrate Open Access champions and identify faculty members to approach, Thomas and Shouse needed to analyze the data by author. The roughly 2,800 unique authors in the dataset had a total of more than 9,100 authorships because of the high number of collaborations. The number of authorships is higher than the total number of articles because duplication is counted when there is a co-author. These multiple-authored papers were fairly concentrated by subject area and extended author networks, though, because nearly 1,500 articles had only a single author. About 190 authors had ten or more articles, accounting for 35 percent of the authorships. The researchers used the Scopus author profile numbers to create author-specific lists used later in outreach activities that included total counts of articles along with the number that were available OA. Two examples include ECU faculty member J. A. McCubrey, who had fifty articles in the data set, forty of which were available OA, and J. G. L. Lee, who had thirty-nine OA articles out of a total of forty-six.

Thomas and Shouse began reaching out to authors while continuing to work with the data. They decided to begin with some of ECU's most prolific authors, depositing available OA articles into The

ScholarShip, ECU's institutional repository (IR), ahead of meeting with the author. These OA articles were generally either gold Open Access or author manuscripts from PubMed Central, along with a few green OA articles from other repositories. When we met with the author, we would describe our project, congratulate them on their publications that were already OA, and offer to deposit permissible copies to the IR on their behalf.

How did the author outreach go? Very slowly–it was a time-consuming process to parse out information by author, schedule, and then conduct any follow-up that was needed. Although some articles were listed multiple times because of co-authorships, several of the authors we spoke to presented a practical solution: they deferred to the corresponding author. We only initiated contact with authors no longer at ECU when they were corresponding authors to whom we were referred by someone still at our university. If an article had no author who was still at East Carolina University, we would not pursue depositing a copy to our IR.

When we examined the data by journal, we realized that our authors' publishing concentration in some titles provides significant Open Access opportunities. There were 2,426 unique titles in the dataset, of which approximately 1,500 had only one article indexed. Thirty-four journals had ten or more articles indexed, and all of these journals offered OA publishing options. The top five journals were all OA: in descending order they are *PLOS One*, *Scientific Reports*, the *North Carolina Medical Journal*, the *International Journal of Environmental Research and Public Health*, and *Oncotarget*.

When analyzed by publishers, we found 261 unique publishers, but a very uneven distribution. The top five publishers were Elsevier, Informa (Taylor & Francis), Springer Nature, Wiley, and SAGE; together they accounted for 61 percent of the articles in our data set. All of the top twenty publishers have Open Access options, and four of them are OA publishers: MDPI (ECU's 8th highest publisher), PLOS (#10), Frontiers (#11), and the North Carolina Institute of Medicine (#20).

With respect to funding status, the largest number of ECU articles were not funded (56 percent). There was a public access requirement (PAR) for 32 percent of the articles, and there was funding without PAR for 12 percent. We have known for some time that the National Institutes of Health (NIH) was ECU's top funder, and it accounts for some 60 percent of all funded articles. Rounding out the rest of our top ten funders in terms of the number of funded publications are the National Science Foundation, the National Natural Science Foundation (China), the United States Department of Agriculture, Centers for Disease Control, Department of Energy, American Heart Association, Department of Defense, National Oceanographic and Atmospheric Administration, and the Natural Sciences and Engineering Research Council of Canada. All of these top ten funders have an OA or public access requirement.

How does ECU's Open Access output compare with peer institutions? Thomas and Shouse took the list of eleven peers identified by the Institutional Research Office and used Scopus to look up their total number of final articles for each of the six years of study, along with the Scopus-identified number of OA articles for each year. The peers were Ball State University, Central Michigan University, Florida Atlantic University, Illinois State University, Kent State University, Northern Arizona University, Ohio University, the University of Nevada at Las Vegas, Utah State University, Washington State University, and Western Michigan University. ECU's OA publishing averages in the high forties compared well overall with those of its peers, which ranged from an outlier low of twenty to a high of fifty-nine. In order to maintain our competitive advantage, however, ECU faculty members will need to continue increasing their OA publishing decisions, since our peer averages seem to be increasing per year. For more information, see Figure 3, ECU Comparison to Peer Lows and Highs.

We were also curious about how we compared to some of our fellow University of North Carolina (UNC) system schools, and repeated the process with data from North Carolina State University, UNC-Chapel Hill, UNC Charlotte, and UNC Greensboro. UNC-CH maintains the

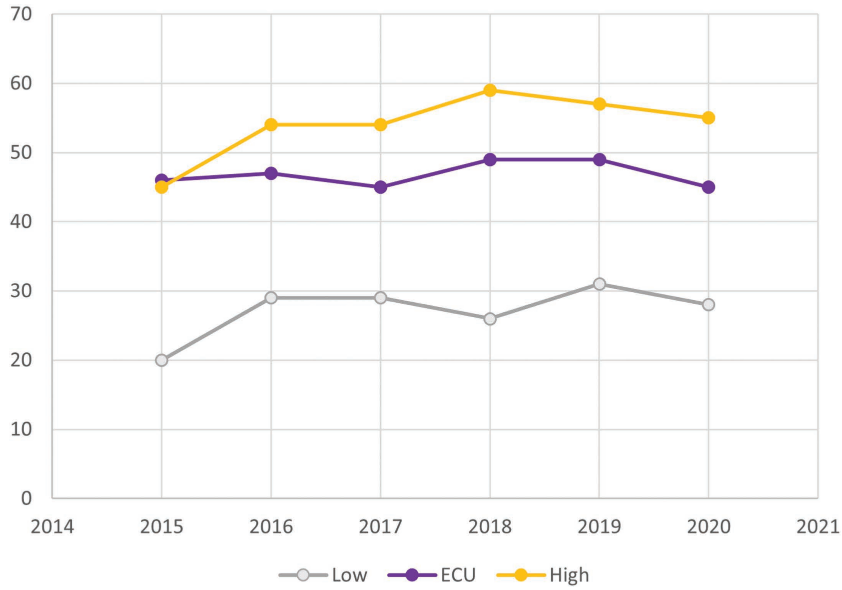

Figure 3. ECU Comparison to Peer Lows and Highs

highest OA averages, ranging from 69 to 73 percent– an advantage due, at least in part, to substantial research funding and an institutional OA mandate.[1] ECU's performance tracks well with NCSU and UNCG, with UNCC close behind and catching up each year.

After establishing ECU's current open access publishing rates and beginning outreach to authors, where do we go from here? One of the most straight-forward actions that we will take is to add publication data for 2021 (and soon 2022). We will continue to reach out to campus authors in an attempt to build a critical mass of recent OA articles and convince as many authors as we can reach of the value of providing OA to their research products. Future actions will also include analyzing publisher distribution to determine whether recent OA agreements have an impact on campus authors' decisions and might also lead us to discussions with the Research Office. One request would be for the Research Office to require either OA publication or deposit of a green OA copy to the IR for all ECU-funded articles. Getting research administrators on board–especially Associate Deans for

Research in the various colleges—we hope will help encourage more faculty members to choose OA whenever possible. It is our hope to change faculty members' attitudes on campus, to choose OA venues when possible, to write publication charges into their grant applications, and to share manuscripts with us in the library when they publish in subscription-access journals that permit it, so we can deposit on their behalf.

Contributor Note

William Joseph Thomas is Assistant Director for Collections and Scholarly Communication for Academic Library Services at East Carolina University, Greenville, North Carolina.

Note

1 "Open Access Policy," University of North Carolina at Chapel Hill, last modified May 13, 2016, https://policies.unc.edu/TDClient/2833/Portal/KB/ArticleDet?ID=132180.