Collaborating to Analyze E-Journal Use Data: A Discussion of Cross-Institutional Cost-Per-Use Analysis Projects within the UNC System

Virginia Bacon
East Carolina University, baconv@ecu.edu

Patrick L. Carr
East Carolina University, carrp@ecu.edu

Virginia Bacon and Patrick L. Carr, "Collaborating to Analyze E-Journal Use Data: A Discussion of Cross-Institutional Cost-Per-Use Analysis Projects within the UNC System" (2012). *Proceedings of the Charleston Library Conference.*
http://dx.doi.org/10.5703/1288284315102

Collaborating to Analyze E-Journal Use Data: A Discussion of Cross-Institutional Cost-Per-Use Analysis Projects within the UNC System

Virginia Bacon, Electronic Resources Access and Discovery Librarian, Joyner Library, East Carolina University
Patrick L. Carr, Head of Electronic and Continuing Resources Acquisitions, East Carolina University

Abstract

This presentation discusses two projects within the University of North Carolina (UNC) system in which the system libraries collaborated to share data to make cross-institutional analyses of expenditures, use, and cost-per-use (CPU). The first project was initiated in 2011 and involved the analysis of e-resources at four UNC libraries. The second project was a UNC system-wide project that occurred in May 2012 and involved comparisons of expenditure and use data for e-journal subscriptions across the system.

Introduction

The transition from print to e-resource collections has created unprecedented potential for libraries to collaborate in the collation and analysis of use data. In this presentation we will consider how libraries can harness this potential to better understand and enhance return on investment (ROI) for their e-journal subscriptions. Specifically, we will discuss two projects within the University of North Carolina (UNC) system in which the system libraries shared data to make cross-institutional analyses of expenditures, use, and cost-per-use (CPU). The first project, initiated in early 2011, centered on the analysis of e-resource CPU data shared among four UNC libraries. The second project was participated in by 15 UNC libraries in May of 2012 and resulted from a mandate issued by the UNC General Administration to compare the expenditures for and use of the libraries' e-journal subscriptions. Throughout the discussion of these projects, we will emphasize the opportunities and challenges of collaborative analysis of e-journal use data.

Harnessing Use Data to Evaluate Collections

Libraries today are well equipped to evaluate their e-resource collections. For example, the COUNTER standard gives libraries a code of practice that e-resource access platforms can adopt to consistently record and exchange use information and then make that information available to libraries. Of course, COUNTER-compliant use data isn't perfect—for example, systematic downloading and provider errors in data collection can sometimes throw off the data—but overall, this data is a very powerful tool for calculating and assessing e-resource use.

In today's information landscape of tight budgets, effectively harnessing use data to evaluate collections is of increasing importance. Libraries are faced with tough choices about how to cope with scarce funding, and a really important tool to help make these choices is CPU. CPU is simply a calculation of an e-resource's annual subscription cost divided by use.

Despite the power of e-resource use data, it is essential to contextualize this data within other forms of collection evaluation. The sociologist William Bruce Cameron once wrote that "Not everything that can be counted counts, and not everything that counts can be counted." The importance of this quote to our current discussion is that use data can never stand alone as a library's only method for evaluating their e-resource subscriptions. Indeed, libraries also need to consider input from students, faculty, and subject librarians. This input can give libraries crucial qualitative information to help make the best decisions.

But when we are aware of its limitations and when utilized in the appropriate context, e-

http://dx.doi.org/10.5703/1288284315102

resource use data is a powerful tool for evaluation. But are we using this data to its fullest potential? When libraries analyze use data, they generally do so in a bubble. But what happens if a library is able to consider use data along with CPU calculations based on the data within the context of data from other libraries?

UNC Cross-Institutional CPU Analysis Pilot Project[1]

These are questions that our colleague Chuck Hamaker considered at a 2010 Charleston Conference presentation. In that presentation, Hamaker, who is a librarian at University of North Carolina (UNC)-Charlotte, took his institution's CPU data and looked at it in the context of CPU data supplied by our institution, East Carolina University (ECU). After participating in this project, we were intrigued by the possibilities of carrying such an analysis further. For a presentation given at the 2011 ALA midwinter meeting—and then subsequently at some other venues during the spring—we decided to build on Hamaker's analysis. We requested CPU data from several other UNC schools with the rationale that the more schools supplying CPU data, the better equipped we are to assess what this data means and how we can use it. We were ultimately able to get two other schools to supply their CPU data, UNC-Greensboro and UNC-Wilmington.

Each of the four participating libraries was asked to enter information concerning CPU for a spreadsheet listing 78 resources. The basis for selecting these particular resources was that they were those that Hamaker had used during his initial research. In other words, we were just building on the data that Hamaker had gathered. Of course, because different libraries subscribe to different resources, there were many resources for which certain of the participating libraries were not able to provide CPU data.

Although time limitations prevent us from a detailed discussion of the project's results,[2] a brief example of the kinds of insights we derived can be culled from one category of the resources that were analyzed, commercial publishers. In comparing the four institutions' CPU for their Elsevier, Emerald, Sage, Springer, and Wiley-Blackwell e-journal subscriptions, we found that the overall average CPU for the publishers across the four institutions was $8.57. Emerald ($3.16), Elsevier ($5.65), and Sage ($6.15) had the three lowest CPU averages, whereas the other publishers all had CPU averages of over ten dollars. As far as institution-by-institution results, we found that ECU and UNC-Greensboro had almost identical CPU averages ($6.59 for ECU and $6.52 for UNC-Greensboro). UNC-Charlotte had a slightly higher CPU average of $7.35, and UNC-Wilmington had a significantly higher CPU average of $13.80.

As we noted earlier, the pilot project's data was presented at several venues in the spring of 2011, and one of the recommendations we made was that, as budgets tighten and as our users' expectations for seamless access continue to grow, libraries must strive to harness their full potential for partnership through the collaborative analysis of e-resource use data. Therefore, we have been advocating that libraries would benefit from building on this project and proactively working together to share cost and use data to make cross-institutional assessments of ROI.

In response to the presentations, we have generally received positive comments about the potential usefulness of such a collaborative project and the opportunities for carrying out such a project in a broader and more systematic way. But these comments were just comments, and it looked like nothing was going to happen to build on the initiative.

UNC System-Wide E-Journal Survey

Background

And nothing did happen until May of 2012. At that time, the UNC system General Administration (GA) actually instituted a project that closely resembled our pilot project but on a larger scale. Although this UNC system-wide project was developed independently from our pilot project, it involved the same basic principles of libraries collaborating to share and analyze e-resource cost and use data to enhance ROI for their e-resource collections.

The project's genesis was a February 2012 request from GA to the North Carolina Office of State Budget and Management's (OSBM). The request was for help in reviewing the UNC system to identify potential efficiencies and cost savings. One component of the review of operations was the UNC system's expenditures for and use of e-journal collections. GA OSBM aimed to discern patterns in ROI and then carry out steps for the system libraries to work collectively to improve ROI.

Data Collection

ECU's work on the project began on May 18, 2012. The deadline for data submission was June 4. To kick off our work on the project, we participated in a conference call with representatives from the various UNC libraries. During the call, we discussed how to gather the data, and we also discussed the survey instrument. This instrument consisted of a spreadsheet with three tables. The first table asked libraries to provide overall expenditures and title counts for all of their journal subscriptions over the 2009, 2010, and 2011 fiscal cycles. The second table asked libraries to provide information regarding 2011 fiscal year expenditures, full-text article downloads, subscription model, and price caps for the 13 publishers that the survey organizers deemed to be of most interest. The third table asked the libraries to provide additional information regarding their subscriptions from those 13 publishers, including pricing, title counts, and full-text article downloads.

Although some of the data collection was centralized though the work of UNC-Greensboro librarians that acted on behalf of the system libraries, the project still involved a lot of work at ECU. This work was carried out by the three librarians at ECU who work primarily with e-resources: Patrick Carr, Virginia Bacon, and Beth Ketterman. The three of us coordinated our efforts to collect and submit all the necessary data effectively and on time.

Recommendations of the UNC Report

Following the June 4 data submission deadline, the accounting firm that GA retained to coordinate the project collated and analyzed the UNC libraries' data. The results of the analysis were presented to GA in an August 2012 report, which aims to provide a "performance baseline" for the libraries, with the primary measure of performance being changes in cost relative to changes in access. The report concludes that, on the whole, UNC libraries are outperforming national averages in the containment of journal price inflation. The report states that e-journal prices have grown nationally by about 9% each year, whereas UNC libraries have limited their e-journal expenditure growth to just 5% each year. Nevertheless, the report indicates that the libraries need to address the problem of declining ROI and price inflation rates that exceed budget growth in higher education. The report presents the following strategies for UNC libraries to lower costs and increase access:

- Creating an online repository that UNC libraries can use to share expenditure and access data;
- Reducing expenditures for high-volume products in libraries through the use of the online procurement company SciQuest (this recommendation has since been dropped);
- Creating a standard template and checklist for e-journal licensing;
- Evaluating and pursuing strategies to encourage the publication of the results of UNC research in Open Access venues;
- Creating a system-wide plan to limit expenditures and increase access.

The report provides the most detail regarding this last recommendation. It indicates that this system-wide plan should focus on four publishers that, based on the data collected, were deemed to be "high risk": Elsevier, Wiley-Blackwell, Taylor and Francis, and Oxford University Press. Additionally, the report indicates that the plan should try to achieve the following objectives by the end of 2015:

- Limit annual increases in expenditures so that it parallels increases in use;
- Limit annual increases in CPU and cost-per-title (CPT) so that it is less than annual increases in expenditures;

- Limit annual changes in CPU or CPT.

Analysis

We believe that the UNC report is an excellent starting point for understanding ROI for the e-journal collections in UNC libraries. But we also believe that the report contains certain flaws, and also that it overlooks certain important implications of the data.

Certain of the flaws in the report are methodological. The report seems to assume that the 13 publishers selected for close data analysis will together provide an accurate picture of the UNC libraries' overall ROI for e-journal collections. However, in some instances the publisher selections seem very questionable, and certain important publishers seem to have been excluded. Another methodological flaw consists in the guidelines for data collection. In certain respects, these guidelines were quite confusing, and as a result the accuracy of some of the survey results is questionable. However, overall we think that the data is a "good enough" picture of the ROI of UNC libraries.

As for the report's five recommendations, we agree with the recommendations concerning collaborative planning, the creation of a shared repository for use data, and the creation of shared licensing guidelines. The other two recommendations, however, we think are less useful and less practical as means for improving ROI in the short term. We think that the UNC libraries should give top priority to the recommendations to create an online repository for sharing expenditure and use data and to create a system-wide plan to limit expenditures and increase access.

While we support the recommendation to develop a system-wide plan, we feel that it is crucial that they not just be limited to the four publishers that the report describes as system-wide "high risk" publishers: Elsevier, Oxford University Press (OUP), Taylor & Francis, and Wiley-Blackwell. In general, we think that making broad, system-wide claims about publisher performances ignores institution-specific context, which is very important. In reviewing the data, we found that there were a number of publishers in which heavy expenditures and use by a small number of institutions lead them to be assessed as high-performing publishers even though they were low-performing at some UNC institutions. Additionally, we found that, when carefully reviewing the data, certain of the "high-risk" publishers fared more favorably when factors such as subscription model and relative CPU were considered and properly contextualized.

The survey also showed an inverse correlation between institution size and CPU: in general, the larger the institution, the lower the CPU. Further review of the data showed that, actually, the correlation is between CPU and research intensiveness: the more intensive an institution is in terms of its research activities, the lower the institution's CPU. This finding suggests that publisher pricing models do not fairly accommodate for research intensiveness.

Conclusion

Overall, the UNC system-wide e-journal survey represented an important step forward. The project drew on the principles of collaboration and partnership demonstrated in the 2011 pilot project of four UNC institutions in order to effectively coordinate the data collection of all UNC libraries and then developed a report that both detailed the findings of their data analysis and presented recommendations based on these findings. Although there are certain aspects of the analysis and certain recommendations that we have called into question, the data collection and report represent a crucial first step in an effort among UNC libraries to share cost and use data maximize ROI for their e-journal collections.

[1] Although this section is written in first-person plural (i.e., "we"), only one of the authors, Patrick Carr, actually participated in the pilot project that is described. The section is written in first-person plural in order to be consistent with the other sections of the write-up.

[2] Details concerning the results of the project are accessible online at http://thescholarship.ecu.edu//handle/10342/3143