The High Cost of "Free" Access: Cost Analysis of Citations Collected from Florida State University

Kelly Grove, *STEM Research and Learning Librarian, Florida State University*

Introduction

Researchers have become accustomed to accessing articles over the internet without having to step foot in the library (Fitzgerald, 2018; Inman, Blevins, Ketterman, & Young, 2019). However, what most do not realize is that academic libraries do their best to make access to online content seamless for researchers both on- and off-campus. The ease of access to content relieves a lot of frustration for both faculty and students as they are able to work from nearly anywhere. Without having to come to the library some researchers have stop seeing their libraries as a partner in the production of research because it is easy to forget their library is paying for institutional access to databases and journals. These factors play into campus researchers and administration questioning the size of the library's budget; as libraries are typically non-revenue producing campus constituents. As a result, libraries feel pressure to demonstrate their value to their own institutions by supplying administrators with data regarding building usage, gate or floor head counts, amount of physical or electronic items, and statistics on consultations, workshops, and classes taught (Kelly, Hamasu, & Jones, 2012, p. 657).

With the objective of providing support for the return on investment of the library; the author conducted a citation analysis with the purpose to recreate bibliographies without access to library networks or resources. The author utilized Google Scholar to locate cited sources, find the cost of access via publishers' websites, and identify any free versions available. The results of this project create a picture of what it would cost a researcher to produce their own published research papers, based on their cited articles if they had to pay for the articles themselves.

Research Question: What does it cost a researcher to produce a paper when they do not have access to an academic library for source material?

Citation Collection Method

The collection of references cited in FSU published papers was completed in three parts.

Part I: The process of collection and sorting of FSU authored papers for cited references is as follows:
1. Using Web of Science Core Collection Database a search with the limiters of: Organization-Enhanced = "Florida State University OR Florida State University," Year Published = "2016," and Document Type = "Article" was conducted.
2. Article bibliographic information was then downloaded by accessing the Create Citation Report feature and sorting the list from First Author -- A to Z.
3. The list was downloaded as Excel Files in groups of 500 (the largest group set allowed in this feature) and combined into one master Excel spreadsheet.

Part II: The process of collection and sorting of cited reference bibliographic information is as follows:
1. Each FSU authored article was searched for by title in the Web of Science Core Collection database.
2. The number of reported cited articles was recorded with the bibliographic information of the FSU authored paper.
3. A .cvs file containing a list of the cited references and their bibliographic information were downloaded.
4. The downloaded .cvs file was converted into an Excel spreadsheet and standardized.
5. Each set of cited references were placed into a master cited reference spreadsheet with a primary key that associated the cited articles back to its parent FSU authored paper.

Part III: The process of creating a sample group of cited references is as follows:
1. 206 FSU papers, representing 10.3% of the FSU authored papers, were randomly selected using a random number generator.*
2. The bibliographic information of the cited references of the selected FSU authored papers were copied into a new Excel sheet.

*The citation sample group was based off of the collection of at least 10% of the FSU authored paper in order to best determine the average cost per article rather than collecting the price of a random 10% of the cited references.

Price Collection Method

To determine the cost to access the cited references Google Scholar was used to find the articles both from the publisher's website and any freely available versions such as institutional repositories or researcher's personal webpage. In order to properly conduct this portion of the study the searches were done while not connected to the campus network or logged into the library proxy server, this ensured that paywalls would come up when trying to access the full text.

1. Using Google Scholar the title of each cited reference in the sample group was searched.
2. The first link for the article was clicked and the cost to access each article was recorded along with important notes in a spreadsheet next to the bibliographic information.*
3. Under the Google Scholar result for the "All [Number] Versions link was clicked to see if free versions of the articles were available.

*Notes included if a free version was found, if the article was open access, if there was a restriction on the length of access to the article, if there were other purchasing options such as renting or cloud access, if the article was international the price in the original currency was recorded, if the cited reference was noted to be a book chapter, or if the reference could not be found or was only a citation in Google Scholar.

Results

The number of cited references reported indicated that the bibliographic information of 99,871 citations should have been downloaded. This would average out to 50 cited references per FSU authored paper. However the bibliographic information of only 74,686 were downloaded. It is assumed that the remaining 25,185 (25%) cited references were papers either not indexed by Web of Science or were incorrectly cited and could not be indexed. Of the sample group it was expected that 10,363 cited references would need to be evaluated, but the bibliographic information of 7,843 cited references could be downloaded from Web of Science; this is keeping in line with missing 25% of cited reference from the larger bibliographic data. During the price collection process the price of 274 (3.5%) articles could not be found either due to the inability to locate the cited reference on Google Scholar or the publisher required a full subscription to the journal or did not allow individuals to access their content. If the average of 50 cited references per paper is used it would be as if 5 and half papers could not be published if researchers relied solely on what could be found and accessed from Google Scholar.

Average Cost to Procude a Paper	**$1,042.03**	
Average Cost per-cited-refernce	**$28.36**	

Research Area	Number of FSU Articles in Sample	Cost to Produce
STEM	129	$141,956.10
Humanities	4	$1,198.65
Social Sciences	40	$39,001.49
STEM/Social Sciences	32	$31,771.95
Social Sciences/ Humanities	1	$730.80
Totals	**206**	**$214,658.99**

Average Cost To Produce An Article

	STEM	Humanities	Social Sciences	STEM/Social Sciences	Social Sciences/Humanities
Cost	$1,100.43	$299.66	$975.04	$992.87	$730.80
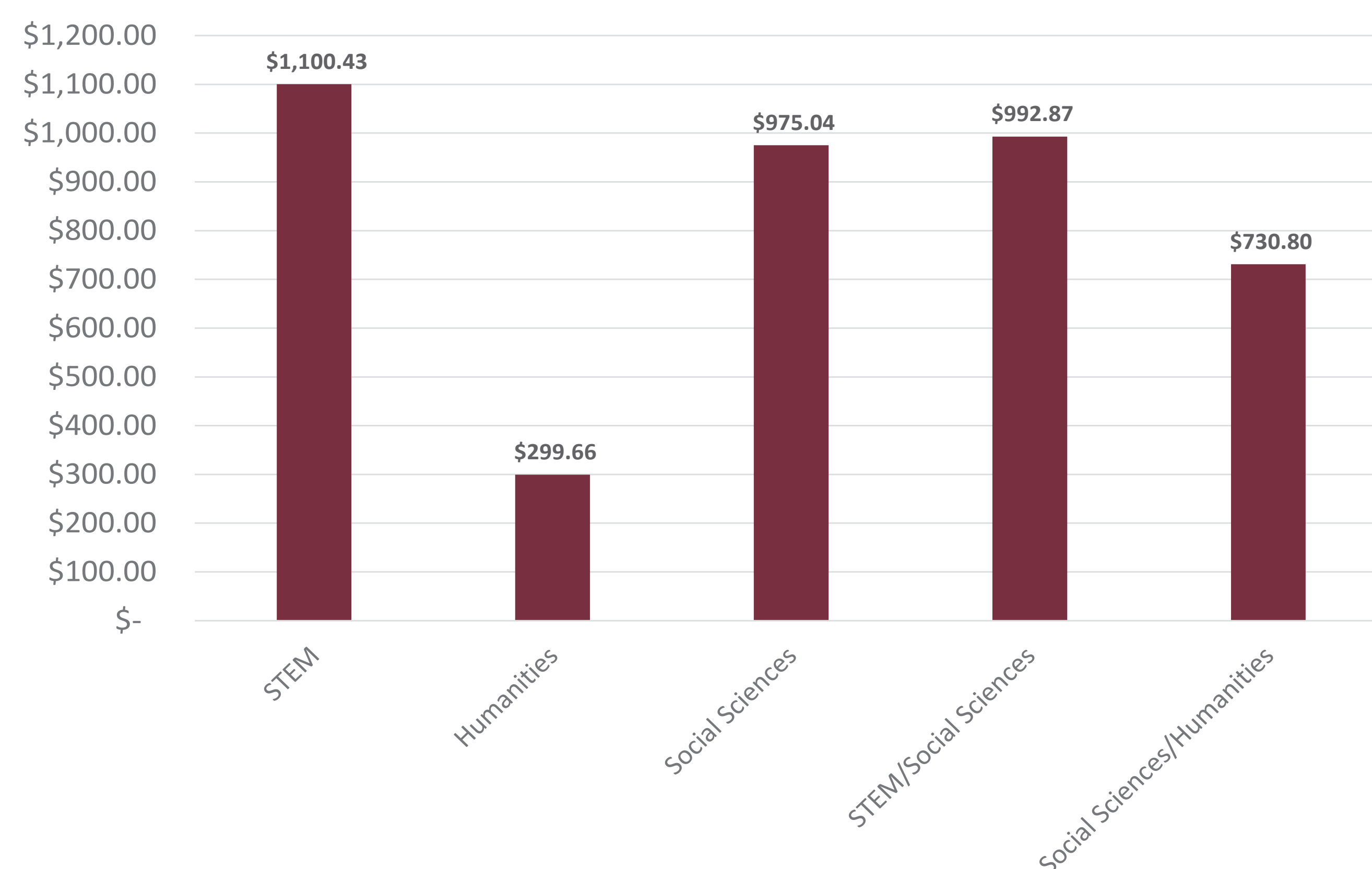
	Number of Articles	Total Cost
Open Access	1487	$0.00
Must Purchase Access	2738	$94,879.91
Free Version Found	3618	$119,696.48
Totals	**7843**	**$214,685.99**

Open Access vs. Purchase vs. Free Version

- Open Access — 19%
- Must Purchase Access — 35%
- Free Version Found — 46%
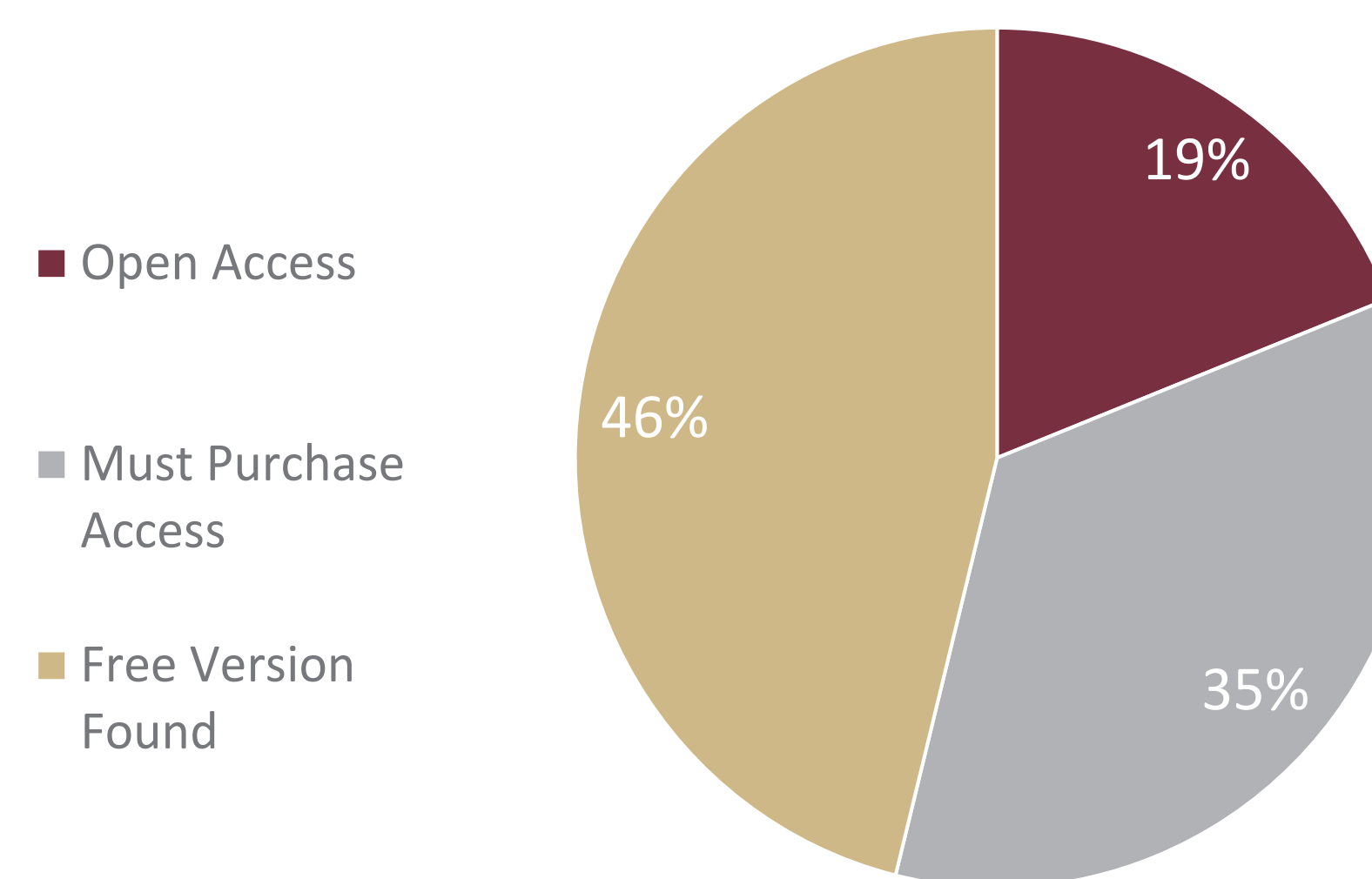
Discussion

Citation Analysis projects are unobtrusive research methods that allow librarians to systematically review the reading, research, and publication habits of a selected group of researchers. These type of projects are usually conducted with the intent to evaluate how the library's collection is meeting the needs of their patrons and what resources are being used the most; however there are some who see citation analysis as tool in exploring the cost per citation (Hoffmann & Doucette, 2012). Additionally citation analysis studies provide data on specific populations at a point in time not available anywhere else. This generated data creates a measurable outcome that can balance out subjective opinions about library collection and use (Black, 2013, p. 293). To the author's knowledge and research no citation analysis has recently or is currently being done at Florida State University to assess the use of library collections or the citation behavior of the institution's research producing population.

Conducting a citation analysis and reporting on the citation habits of FSU researchers and using that data to talk about how the library's collection is or is not meeting the needs of researchers does not grab the immediate attention of stakeholders. It is just another report of numbers and statistics. To capture administration's attention attaching dollar amounts to that data allows the library to demonstrate its value.

The biggest hurdle that had to be overcome in the project was time commitment. The batch downloading of FSU authored papers was straight forward, but when it came time to download cited references the workflow became significantly slower, as each paper had to be looked up individually in order to pull the associated references. However, the largest time commitment came when it was time to collect prices and look for possibly freely available copies of each cited reference in the sample group. This work could not easily be done while with in the FSU Wi-Fi network. Working from home and after "normal" work hours became a norm for completing this portion of the data collection. With the stumbling block of time came the inability to work consistently on this project due to changes in work responsibilities which caused temporary rearranging of priorities.

Additionally the initial method of searching and recording the results of searches from Google Scholar required some trial and error. Some publisher websites made it difficult to find the price of a onetime purchase either by requiring the user to go through a long series of clicks or by requiring them to create a login for their site. At times this would considerably slow down the searching process. Developing a coding system to use could not be fully done until the process was started as it would have been impossible to anticipate all of the possibilities of access options.

Moving Forward

This citation analysis case study supplies a snapshot of research interests and publication output from Florida State University during the 2016 calendar year. As we are now in 2019 there is an argument that the prices of articles collected are directly applicable in open access, made freely available and ones that must be purchased could be very different. Time will always change the numbers, but by starting to look at the past will allow us to move forward as time marches on. Ideally this snapshot is just a starting point in the exploration of how libraries provide a large return on investment for their university. This study and future studies like this can be used in budget talks, inform collection development, be a tool in library services promotions, and be used in cases for promoting open access and institutional repositories.

Dollar amounts speak large volumes to those at the administrative level. By repositioning the view that the library is an overhead cost to be viewed as a money saving tool will place the library in a position to fight for more funding. The next step in this study is to conduct a return on investment comparison. It would be at this time that the cost-per-cited-reference would be compared to the library's collection at during the 2015 and 2016 calendar year. In combination with what the library pays for journal/database access and available use data. The combination of internal library usage data and budget will allow for a more direct comparison to the prices found after search Google Scholar and produce a return on investment report library administration can use in budget meetings with other campus administrators.

For more immediate promotion of the library's value a marketing campaign directed to researchers themselves can use this citation analysis data to create info graphics highlighting the cost of single article purchases based on journal or discipline. This can raise researcher's awareness of what it would cost themselves if they had to purchase access to the papers they need. Additionally by demonstrating the savings produced articles found freely available from institutional repositories should be an incentive for researchers to place their research products into their own institutional

Citations

Black, S. (2013). Practical Applications of Do-It-Yourself Citation Analysis. The Serials Librarian, 64(1-4), 285–298. https://doi.org/10.1080/0361526X.2013.760420

Fitzgerald, S. R. (2018). Serving a Fragmented Field: Information Seeking in Higher Education. Journal of Academic Librarianship, 44(3), 337–342. https://doi.org/10.1016/j.acalib.2018.03.007

Hoffmann, K., & Doucette, L. (2012). A Review of Citation Analysis Methodologies for Collection Management | Hoffmann | College & Research Libraries. https://doi.org/10.5860/crl-254

Inman, M., Blevins, A. E., Ketterman, E., & Young, K. L. (2019). Now Tell Us What You Want: Information-Seeking Habits of Health Sciences Faculty. Medical Reference Services Quarterly, 38(2), 131–142. https://doi.org/10.1080/02763869.2019.1588046

Jabaily, M. J. (2018). An Analysis of a Pay-Per-View Article Token Report: Lessons for Better Understanding Value and Reporting. Serials Review, 0(0), 1–9. https://doi.org/10.1080/00987913.2018.1471651

Kelly, B., Hamasu, C., & Jones, B. (2012). Applying Return on Investment (ROI) in Libraries. Journal of Library Administration, 52(8), 656–671. https://doi.org/10.1080/01930826.2012.747383

Martin, V., Gray, T., Kilb, M., & Minchew, T. (2016). Analyzing Consortial "Big Deals" via a Cost-Per-Cited-Reference (CPCR) Metric. Serials Review, 42(4), 293–305. https://doi.org/10.1080/00987913.2016.1248218

FLORIDA STATE UNIVERSITY

FLORIDA STATE UNIVERSITY LIBRARIES