The effect of discovery systems on online journal usage: a longitudinal study

Updated from a paper presented at the 37th UKSG Annual Conference, Harrogate, April 2014

Many academic libraries are implementing discovery services as a way of giving their users a single comprehensive search option for all library resources. These tools are designed to change the research experience, yet very few studies have investigated the impact of discovery service implementation. This study examines one aspect of that impact by asking whether usage of publisher-hosted journal content changes after implementation of a discovery tool. Libraries that have begun using the four major discovery services have seen an increase in usage of this content, suggesting that for this particular type of material, discovery services have a positive impact on use. Though all discovery services significantly increased usage relative to a no discovery service control group, some had a greater impact than others, and there was extensive variation in usage change among libraries using the same service. Future phases of this study will look at other types of content.

This research study addresses whether the implementation of a discovery service impacts usage of publisher-hosted journal content. A great many academic libraries around the world have implemented discovery services in the past half-decade as a way of helping their users find information in a way that makes more sense to people who have been raised on single search box technology like Google. Searching across multiple resources at once makes it much easier to find information, especially for students who might not understand the difference between an article index and a library catalog, and even for more experienced researchers who may not know the best subject-specific resource for locating articles. With the growth of interdisciplinary research, discovery services make it easier to gather information from a variety of disciplinary viewpoints at once. Usage of discovery systems has forced a paradigm shift in how our users find information, and yet we have not looked very carefully at the impact of that shift on the resulting content they are finding and using. With the notable exception of the recent UKSG/Jisc study[1], prior studies of the impact of discovery system implementation have either used only descriptive statistics or anecdotal evidence and have focused entirely on the experience of an individual institution. This study assesses the impact of discovery systems on a large scale: across multiple institutions, multiple publishers, and multiple discovery services.

In this first phase of a multi-part study, we investigate the impact these services have on publisher-hosted journal content only. Obviously, the information at the heart of this phase of the study is one small part of all information accessible through discovery systems and available to our students and faculty. Furthermore, of this subset, the sample we built for this phase of the study is only six publishers, albeit six significant publishers of academic information, including several very large ones (see Figure 1). To date, we have not included usage of the same journals through aggregated databases or packages, and we have not explored discovery system effects on other types of information, such as e-book or print book or print journal usage. In addition, this phase of the study explores *whether* discovery services have an effect on the usage of online journals, not *why* any effects might exist. After gaining a better understanding of the nature of these effects, we plan to address some of their potential causes with further research.

MICHAEL LEVINE-CLARK
Associate Dean for Scholarly Communication and Collections Services University of Denver Libraries

JOHN MCDONALD
Associate Dean for Collections University of Southern California

JASON S PRICE
Program Manager SCELC

Publisher-hosted journals are only part of the picture

eBooks, pBooks, newspaper articles, aggregator journal content, etc.
publisher journal content
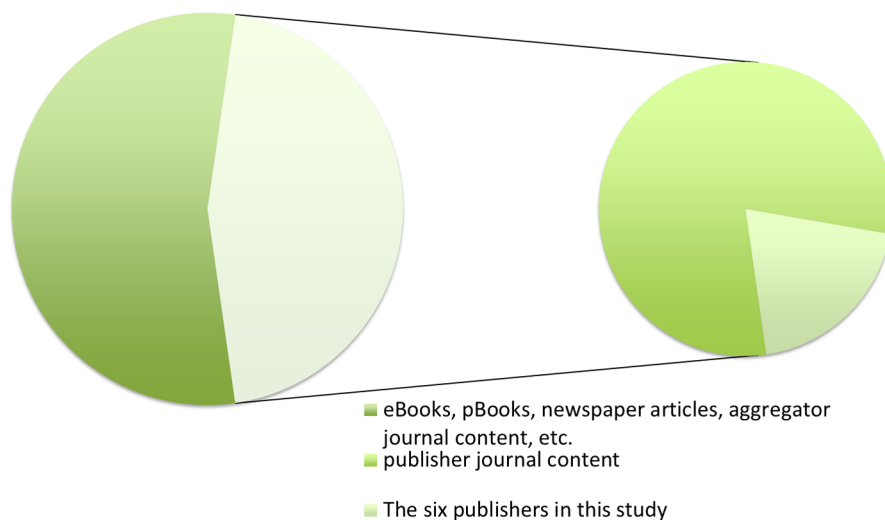
The six publishers in this study

Figure 1. Publisher-hosted journal content only used in this first study

A few assumptions underlie our study: during the period of the study, we assume that most institutions retained a relatively stable user base. This in turn allows us to assume that the total search effort should stay roughly the same year after year. It is obviously going to fluctuate a little bit, but it is not going to rise or fall dramatically unless the user population or perhaps the curriculum changes in some radical way, neither of which is likely at most institutions. Discovery services should draw users away from other search tools, such as the abstracting and indexing tools and the OPAC, which alters the overall productivity of searches, and therefore researchers are going to find different types of information. Based on this, we assume that discovery services have some sort of effect on what people find and how.

Our study analyzes the four major web-scale discovery services, including EBSCO Discovery Service (EDS), ExLibris Primo, Serials Solutions (now ProQuest) Summon, and OCLC's WorldCat Local (WCL). The data set consists of 33 institutions, six each using one of the four discovery tools, and an additional nine libraries serving as a no discovery system control group allowing us to account for background growth in journal usage over time. We contacted each of these libraries and gathered information including the implementation date (the month and the year), location of the search box, and level of marketing or PR about the new service. Twenty-eight libraries from the US were included, two from Canada, and one each from the UK, Australia and New Zealand. The institutions ranged from large ARL libraries to smaller state-supported institutions or liberal arts colleges. The average library size was 1.1 million volumes, as determined by WorldCat holdings, ranging from a high of 2.6 million to just under 300,000.

All of the Primo, Summon and EDS libraries implemented their discovery service in 2011 while the libraries using WorldCat implemented it between 2010 and 2012. The study includes 9,206 unique journal titles from those six publishers held at one or more of the 33 institutions. Our final number of usable library-journal usage change combinations included over 163,000 observations. Library-journal usage change observations that lacked a full 24 months of usage in the COUNTER reports were excluded, as were observations that fell three or more standard deviations above or below the mean of the data set. This treatment of outliers is standard in statistics and resulted in only 332 observations that had a full range of 24 months of usage being removed, or 0.2% of all observations. These outliers may be explained by usage security breaches, systematic downloading of journal articles, and/or usage recording glitches, which are consistent with the observation that they tended to be more common in some institutions (and institution-publisher combinations) than others.

We used COUNTER JR1 reports to compare full-text downloads for the 12 months before implementation with the 12 months after the implementation date. Our dependent variable was the net change in usage from 12 months before to 12 months after.

Each discovery service was represented by an equal number of institutions (n=6), with Primo representing the largest group of observations at 32% and WorldCat Local having about 20%, while EDS and Summon had about 25% each. The data set included three very large publishers (84% of all observations), two medium-sized publishers (14% of the observations) and one very small publisher (only 2% of all observations). (See Figure 2, where the percentage for each publisher is shown.)
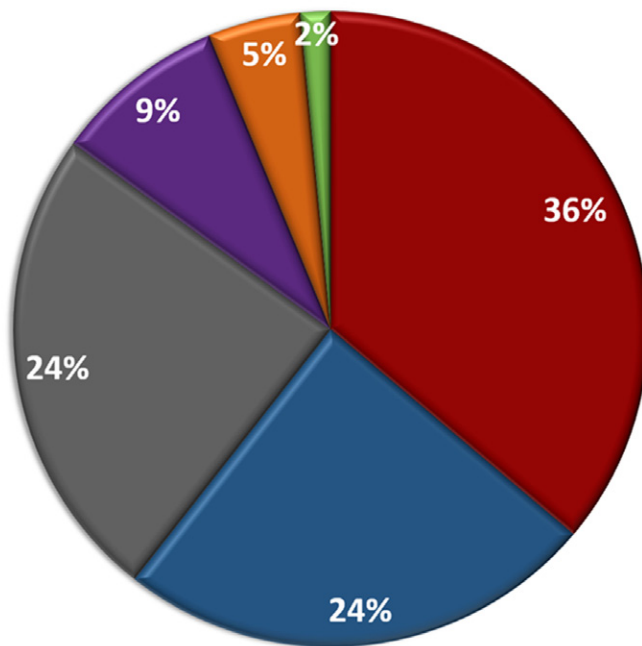
Figure 2. Observations by publisher

Journals grouped by library and discovery service (see Figure 3) show that Primo libraries have the largest average number of journal observations, with Summon having libraries with relatively large collections (as represented by journal observations) using that service, followed by EDS and then by WorldCat Local. The control group included a set of libraries holding a high average number of journals from these publishers, comparable to the Summon libraries on average.
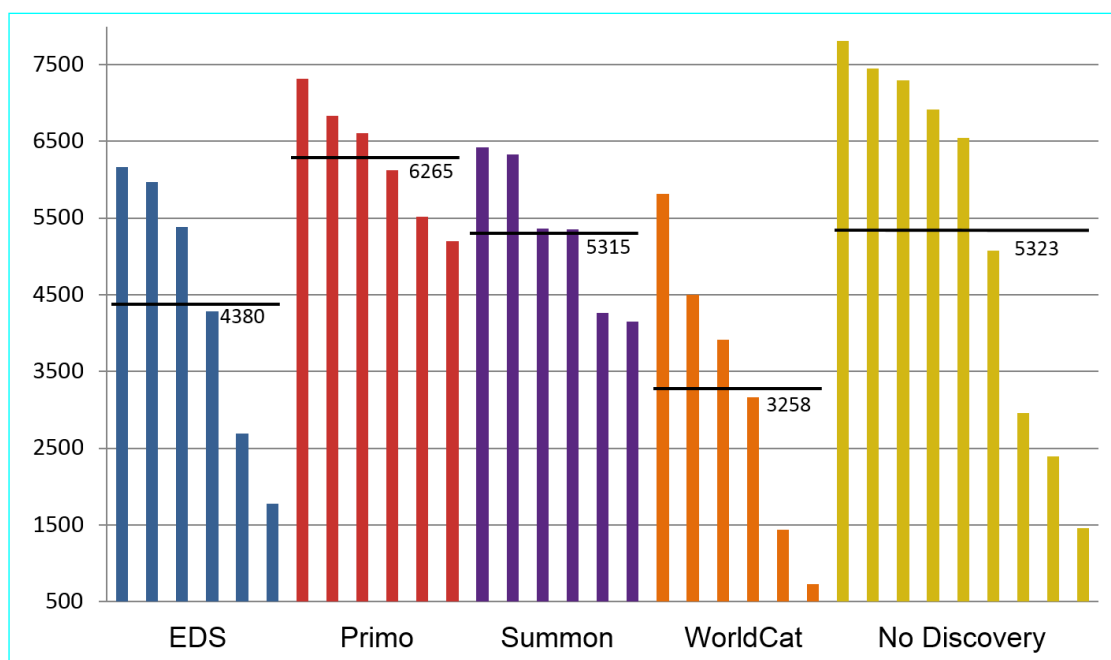

Figure 3. Journals by library and service

Figure 4 represents the libraries ranked by average usage per journals, both before (blue) and after (red) implementation. Only three institutions (highlighted by boxes) had decreases in usage, while 30 experienced increases in usage.
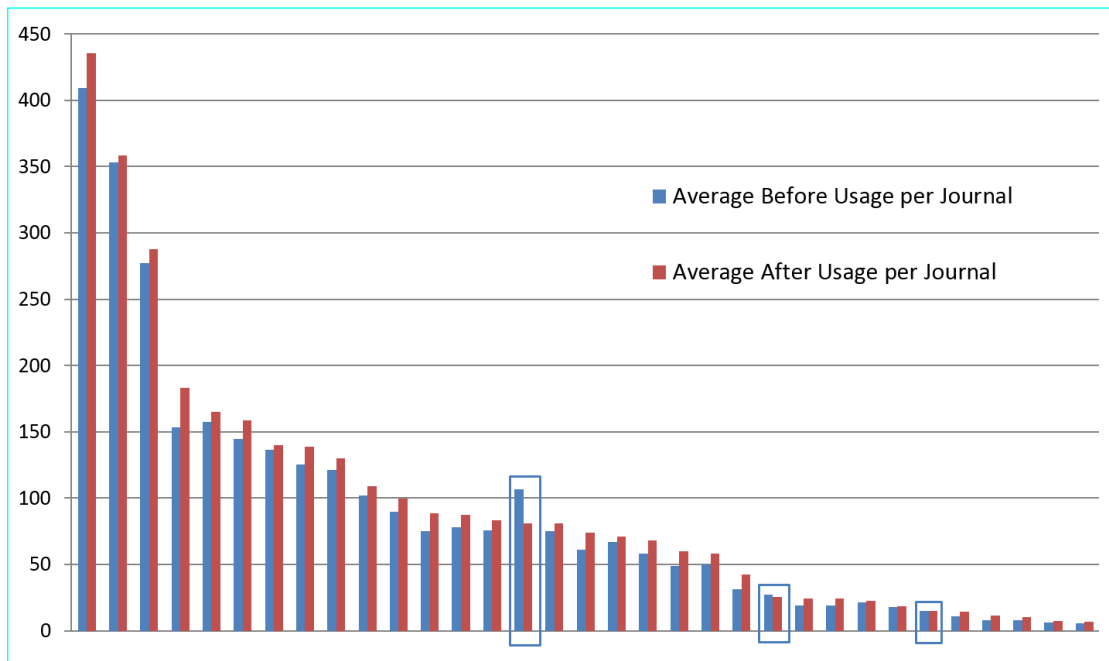

Figure 4. Average journal usage by library

Figure 5 shows each of the six publishers, grouped by discovery service and the average usage change per journal across all libraries using that service. An increase of 0.5 per journal per 1,000 FTE means that an average institution in the study of 10,000 students would have seen its usage of a publisher of 500 journals increase by 2,500 downloads after implementation of that discovery tool. Although most publishers saw increases in their average journal usage, some saw a decrease. Every discovery tool showed extensive variation in its effect on average usage change across the six publishers.
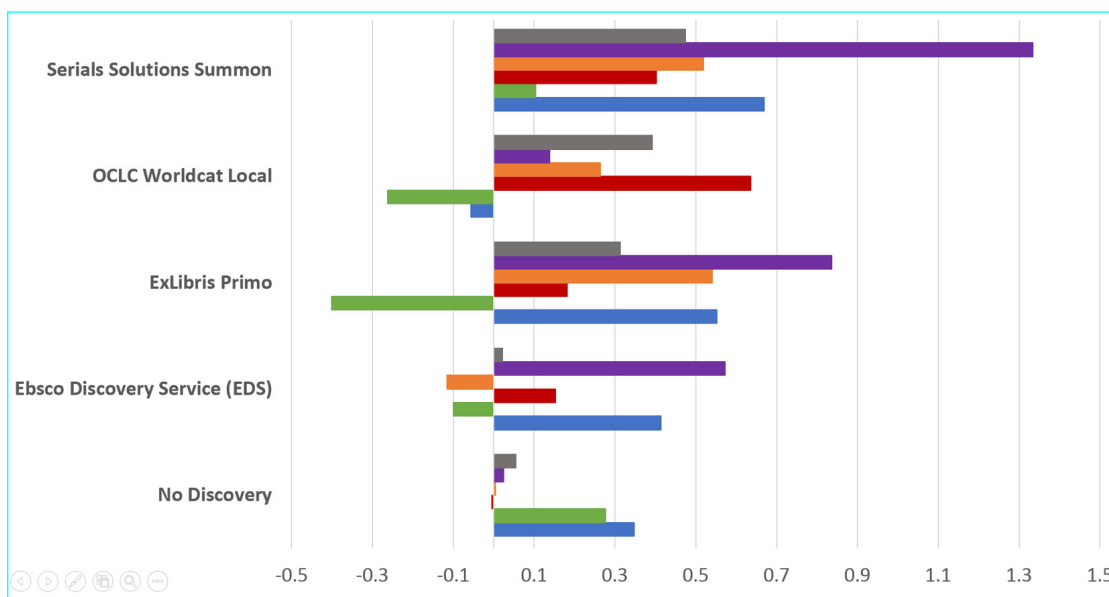

Figure 5. Average usage change per journal per 1,000 FTE, by publisher and discovery service

We found large variations for every institution using these discovery tools, and major variations by publisher within the discovery tools. Some publishers saw net increases for some discovery tools whereas others experienced decreases in usage.

The goals of our inferential statistics were to determine whether those observed differences were significant or resulted from chance effects and to determine which of the three factors – library, publisher, discovery service, or even a combination – contribute to the differences in usage change at the journal level. Our experimental design used a partially nested ANOVA with five factors: discovery service, publisher, the interaction of discovery service and publisher, library nested within discovery service, and the interaction of library nested within discovery service with publisher (see Figure 6).

"We found large variations for every institution … and major variations by publisher"

Factor	df	SS	MS	F	P
Discovery	4	9226	2306.61	20.81	<.0001
Publisher	5	2814	562.87	5.08	.0001
Discovery x Publisher	20	8821	441.07	3.98	<.0001
Institution(Discovery)	28	46673	1666.89	15.04	<.0001
Institution(Discovery) x Publisher	134	89403	667.19	6.02	<.0001
Error	163353	18102576	110.82	--	

Figure 6. Nested ANOVA model results

Due to the complexity of this model, our initial analysis using the error terms supplied by default by SPSS 21 (and presented at the UKSG 2014 conference) did not match the statistical model. Subsequent work with a statistics consultant, completed between the conference presentation and the completion of this proceedings paper, resulted in minor changes to the significance tests for each factor. Theses change did NOT affect the mean values and standard errors displayed in our charts, which are identical to those shown at the conference. Furthermore, the only substantive change to the statistical significance of the results was that the effect of publisher, which was not significant in the initial analysis, was determined to be significant in the final analysis.

The response variable we used to quantify change was change per 1,000 student FTE (full-time equivalents). The advantage of this measure over raw usage change is that it accounts for lower usage (and therefore lower usage change) at smaller institutions.

In response to our first question, whether usage changes vary across libraries (Figure 7), we see the 33 libraries sorted by mean change across the x-axis and the mean change per 1,000 FTE + and -2 standard errors on the y-axis. The standard error bars are a measure of the variability around each mean. Means with overlapping error bars are unlikely to be different from each other, whereas those with non-overlapping error bars are more likely to be different. These data do suggest that there were some differences among libraries in their mean change. This initial view is oversimplified, however, because it doesn't associate libraries with the discovery tool they implemented. Put another way, it does not take into account the fact that libraries were nested within discovery services.

Our second question is whether usage change varies across libraries using the same discovery service. We did indeed find significant differences in usage change among libraries nested within discovery services (df = 28, F = 15.04, p < 0.0001, Figure 8).
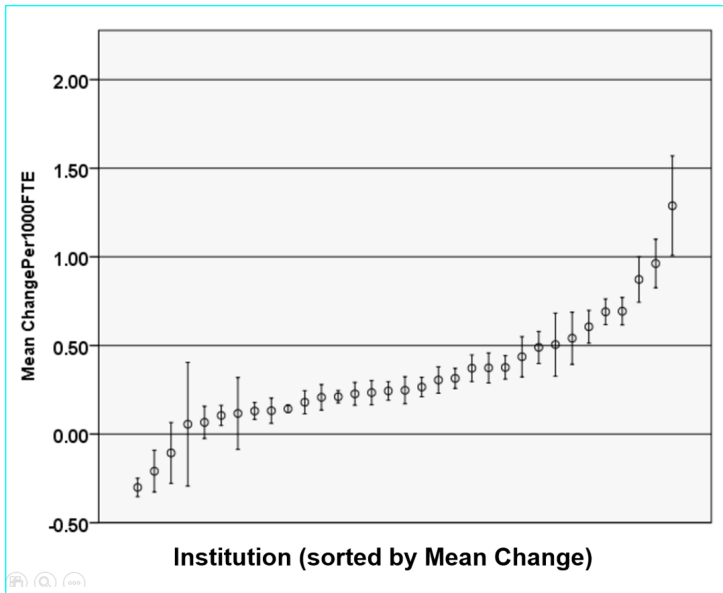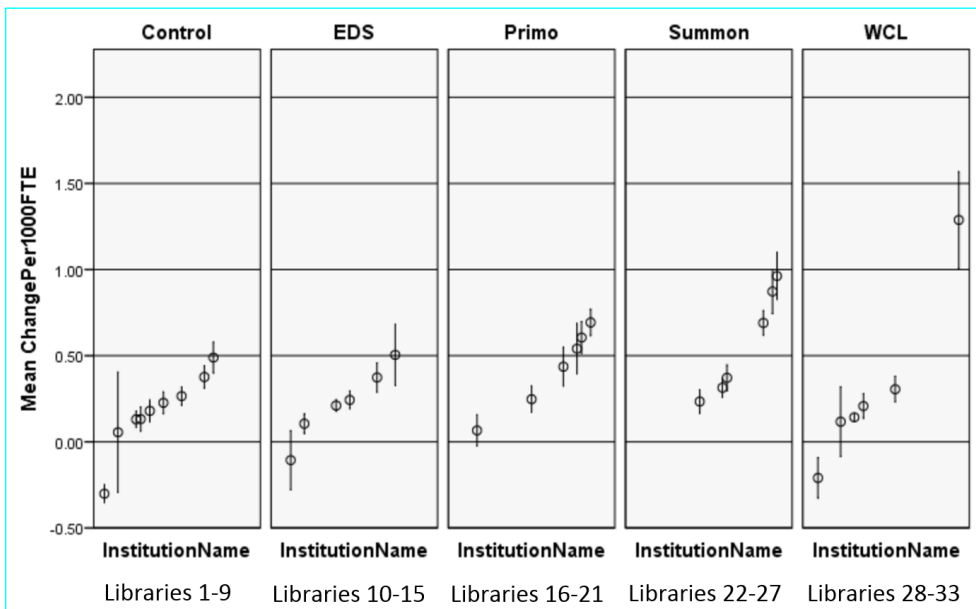
Figure 7. Does usage change vary across libraries?

Figure 8. How does usage change differ across libraries using the same service?

Our third question is whether usage change differs across publishers (Figure 9). We did find significant differences among publishers in their degree of usage change (df = 5, F = 5.08, p = 0.0001), although this effect appears to be weaker than the effect of discovery service based on its lower F value as tested over the same error term. One publisher had a mean change that was not significantly different from zero, three others clustered around a mean increase of 0.23 uses per journal per 1,000 FTE, and two others are about double that value at 0.41 and 0.55. Pairs of means that were significantly different from each other are marked by different letters (Tukey multiple comparisons, p < 0.01).

Our fourth question asks whether usage change varies across discovery services (Figure 10). This is the most fundamental question in our study, and we did find a significant difference in journal-level usage change among samples of libraries using different discovery services (df = 4, F = 20.81, p < 0.0001). Journal usage at Primo and Summon institutions increased to a greater degree than usage at the EDS and WCL institutions we tested, and the usage increases at EDS and WCL institutions were slightly higher than average journal usage increase for the control group of libraries that had not implemented a discovery tool. Pairs of means that were significantly different from each other are marked by different letters (Tukey multiple comparisons, p < 0.05).
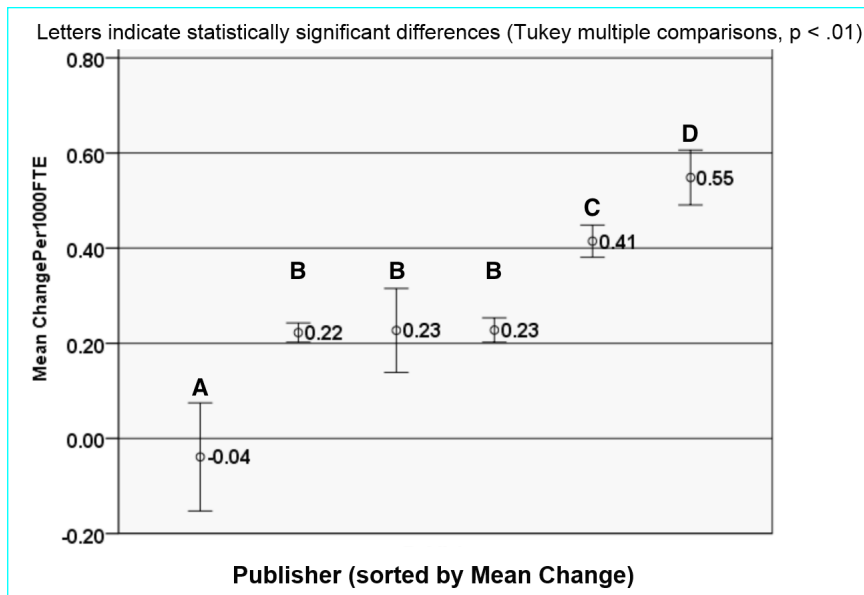
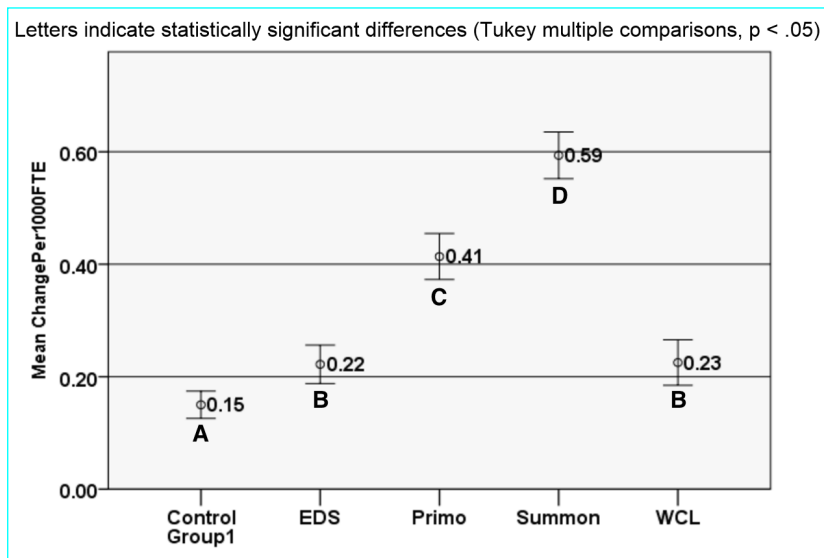Figure 9. How does usage change differ across publishers?

Figure 10. How does usage change differ across discovery services?

The final question we addressed is whether the effect of discovery service differed across publishers (Figure 11). We can think of this question as asking whether the lines connecting the mean usage change of the six publishers for each discovery treatment differ in shape. The effect of discovery service did indeed differ across publishers (df = 20, F = 3.98, p < 0.0001): at least one of the lines differs significantly in shape from one of the others. It is interesting to note that the mean usage change +/- 2 SE was less than or equal to zero for at least one publisher's journals for every discovery system treatment.

In summary, discovery service was the strongest statistically significant variable predicting change in journal usage after implementation. Summon and Primo increased publisher-hosted journal article usage more than EDS and WorldCat Local, which in turn increased usage slightly compared to the 'no discovery' control group.

Library nested within discovery service was the next strongest predictor, and publisher and the interaction of publisher and discovery service also had detectable effects. Summon was the only discovery service whose implementation increased usage across all publishers' content, and there is significant variation in usage change among

"Discovery service was the strongest statistically significant variable predicting change in journal usage …"
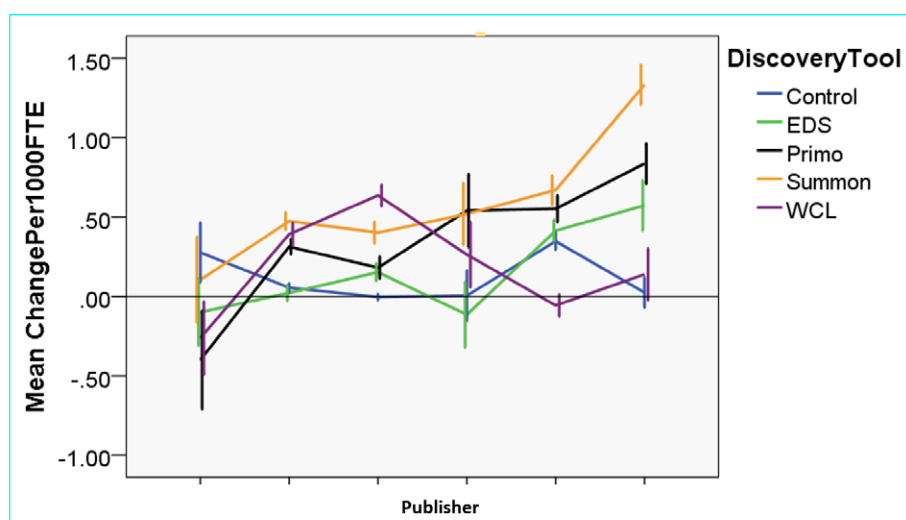
Figure 11. Does the effect of discovery service differ across publishers?

libraries representing each discovery system treatment. There were variations across the board, indicating that each library should examine its own implementation of a discovery service and keep alert to significant increases or decreases in usage of publisher-hosted journal content, since each publisher's content is affected differently by each discovery service. Our results indicate that it would be prudent for each publisher to ensure they are working with each discovery service as effectively as possible to maximize usage of their content, and that discovery service vendors should acknowledge and address the differential effects they have on these journal content providers.

Future phases of this study will incorporate aggregator-hosted journal content, publisher-hosted journal backfile content and e-books. Once we have added in these additional data, we plan to work with publishers, libraries and discovery system vendors to examine implementation options and decisions to help determine how these choices may impact usage.

"… it would be prudent for each publisher to ensure they are working [effectively] with each discovery service."

Reference

1. Spezi, V et al, *Impact of Library Discovery Technologies: A Report for UKSG.* UKSG and Loughborough University, 2013: http://www.uksg.org/sites/uksg.org/files/UKSG_final_report_16_12_13_by_LISU.pdf (accessed 12 May 2014).

Corresponding author:
Michael Levine-Clark
Professor / Associate Dean for Scholarly Communication and Collections Services
University of Denver Libraries
Tel: 011 303 871 3413 │ E-mail: michael.levine-clark@du.edu

ORCID iDs: Michael Levine-Clark: http://orcid.org/0000-0003-3001-0261│ John Mcdonald: http://orcid.org/0000-0003-4499-6295 │ Jason S Price: http://orcid.org/0000-0002-7421-4955