Trends in Digital Newspaper Discovery:

A Natural Experiment

Heather Cribbs [a], Gabriel J. Gardner [b*], Katherine Holvoet [c]

Author Note

[a] Library Systems & Electronic Resource Management Coordinator, Walter W. Stiern Library, California State University Bakersfield, Bakersfield, United States of America; ORCID https://orcid.org/0000-0001-7526-9037

[b] Associate Librarian & Discovery Coordinator, University Library, California State University Long Beach, Long Beach, United States of America; ORCID https://orcid.org/0000-0002-9996-5587

gabriel.gardner@csulb.edu

[c] Electronic Resources Librarian & Assistant Head of Content Organization and Management, SDSU Library, San Diego State University, San Diego, United States of America; ORCID https://orcid.org/0000-0002-0145-8377

Abstract

In this article, the discovery and use of digital newspaper collections are explored by capitalizing on a natural experiment that arose when five California State University libraries activated the Primo newspapers search interface, and five other libraries with similar enrollment numbers and comparable demographic profiles did not. By analyzing Primo Analytics data, COUNTER R4 data, and A-Z database list click-through data collected from the ten libraries over the course of academic years (AY) 2018-2019 (pre-deployment) and AY 2019-2020 (post-deployment), the effects on usage and legible user behavior of introducing a specialized Primo newspaper scope are calculated. Researchers explore how this research method can be a model for libraries to investigate trends within their own organizations.

Keywords: Digital newspapers, discovery, Primo, natural experiment framework, COUNTER

Introduction

Newspaper content is an element of aggregated search results. Discovery layers frequently have two ways of handling newspaper content, having newspaper results integrated into basic and advanced searches and segregating newspaper results into a separate search interface or scope. Primo as a discovery layer was developed by Ex Libris in 2005 and launched in 2010. In the May 2018 Primo Release Notes Ex Libris announced a separate Newspaper Search Interface that was rolled into production environments on June 4th, 2018 (Natan, 2018). Ex Libris claimed "the new feature increased the ability to discover content from newspapers, magazines, and other news resources" with the rationale to increase "focus on scholarly content" within the Primo Central Index. A short configuration guide as well as a FAQ documentation soon followed (Ex Libris, 2018a,   2018b).

Researchers at several California State University (CSU) campuses collaborated to capitalize on a natural experiment that arose when the CSU consortium made implementing a new Primo Newspaper Search Interface optional. After an initial environmental scan, five CSU libraries that implemented the newspaper search interface were identified and paired with five corresponding CSU libraries of similar size, FTE enrollment, and demographic profile that did not. Analysts calculated the effects on usage and user behavior of introducing the Primo newspaper search scope by analyzing Primo Analytics data, COUNTER R4 data, and A-Z database list click-through data collected pre-deployment in AY 2018-2019 and post deployment in AY 2019-2020.

Institutional Context & Sample Data

The California State University originated in 1960 as a result of the California Master Plan for Higher Education. Students are drawn from the top third of the state's high school graduates, and while many CSU campuses offer some Master's and Doctoral level degrees, the CSU is California's primary undergraduate teaching institution. The California State University system comprises 23 campuses located throughout the state, and educates 482,000 students annually, and awards 127,000 degrees annually (Office of Public Affairs 2020).

While the CSU system supports one of the most diverse student populations, there remains a shared mission and vision as well as structure to all CSU campuses. With the implementation of the Unified Library Management System (ULMS) as well as a shared catalog, this central and consortia model continues to unify their campuses. Through this shared catalog the CSU library system has one of the largest centrally managed electronic collections, known as Electronic Core Collection (ECC). The ECC represents the core subject areas and disciplines shared in common by all CSU campuses. CSU campuses have many differences, but there are several unifying characteristics that make comparisons and generalizations as well as in depth study possible. CSU campuses have a baseline curriculum set by the Chancellor's Office, and through the ECC are able to share library resources and access throughout the system.

There were campus level variations in gender split and other demographic qualities such as Hispanic and veteran populations among the campuses selected. Within the study sample, the two largest colleges by enrollment are Business and Engineering. While individual local institutional subscriptions vary, ProQuest Global Newsstream and ProQuest Ethnic Newswatch are shared centrally and available to every campus through the shared consortia catalog. Discovery and access to these electronic collections is provided through two main access points available via each

Library's website, Springshare's A-Z Database list and OneSearch, the CSU branded "single search" Primo Ex Libris interface.

Objectives

As researchers and librarians, the authors are constantly seeking to both better understand their user's behavior, and to improve the search experience. The objectives for this study were:

1. To develop a better understanding of how users seek and use digital newspapers.

2. To evaluate CSU Libraries electronic newspaper database subscription use.

3. And to provide a framework for evaluating the effectiveness of Primo "enhancements" and new product release features.

Literature Review

Several topics in the literature are of specific relevance to this present study, including how users seek and use digital newspapers. This review brings together a body of literature with diverse methodologies and strategies that investigate user behavior in information systems and discovery interfaces as well as studies examining usage metrics and web analytics to better understand user searching behavior.

There is extensive literature exploring user experience with discovery layer interfaces and information retrieval systems. Most research focuses on qualitative data through surveys, questionnaires, and/or focus groups to gather information on user preferences. Yoon used a technology acceptance model (TAM) to understand user attitudes including perceived usefulness

and ease of use of a mobile library application (Yoon, 2016). Traditionally this model is used to examine personal behavior and satisfaction levels surrounding digital library environments and has expanded to include discovery tool features. This survey method study is focused on the perspectives of Librarians and Information Professionals proposing a framework for evaluating information technology. Daramola uses a questionnaire to study perceptions of undergraduates using electronic resources at an academic university in Nigeria and explores the evolution of digital content and subsequent impact on how users search (Daramola, 2016). This research maintains that searchers tend to use only what is readily available and easy to use. Of interest to this present study was the finding that "e-newspapers occupied the third position" in rank of most used resources in the library after e-journals and e-mail and before e-books. Njeze's research targeted newspaper and magazine use within academic libraries through a user survey, mentioning notable literature surrounding newspapers' importance to users for following rapid developments and current trends as well as historical for research purposes (Njeze, 2013).

More recently, Meyer explored the current impact and importance of newspapers on scholarly research through analyzing citation rates of publications referencing newspapers as sources relating to four major newspaper titles (Meyer, 2018). This author's key findings highlight newspapers as a critical part of the academic publishing landscape as use continues to rise. Njeze also notes that interest is not limited to the humanities, but includes many subject fields and disciplines.

Pacy used vendor provided data for two basic metrics, sessions and pageviews/article retrieval to study access for the digital component of a popular, local newspaper within a public library over

a six week period (Pacy, 2014). This research directly highlights the lack of newspaper use studies within the United States. Gooding noted the gap in recent literature surrounding user information seeking behavior for digitized newspaper collections in his research study, which employed "webometric techniques" to evaluate the use of the then newly launched Welsh Newspapers Online discovery interface (Gooding, 2016). While this method explored web server content logs, this research remains highly relevant to this present study. Gooding refers to patterns of engagement and user behavior to draw conclusions in relation to existing information behavior and how a "larger longitudinal data set would increase the study's significance". Xie & Wolfram drew attention to how longitudinal studies have often been applied to analyze digital library environments for insight into user behavior and use patterns (Wolfram & Xie, 2009). These researchers provide a general overview of the extensive research surrounding web search engines and web page searching behavior, and make the distinction between digital library environments and these search interfaces. The findings of this study become less applicable as discovery layers increasingly mimic current major web search engines. While this present study was unable to use a longitudinal method due to the regency of the Primo Newspaper Search Interface, these authors used a similar analysis method as they investigated transaction log summaries to determine the use of ProQuest databases over a three-year period.

Usage data research focuses mostly on subscription journal content, not newspapers. Most usage studies center on the need for better standards or comparison of locally collected data to vendor provided data. Very few studies use analytics to assess how users interact with digital resources beyond cost per use or impact factor. Atanassova brings together discovery concerns associated specifically with newspapers through performing usability testing on a new browsing tool

(Atanassova, 2014). Most research on current discovery tool evaluation involves usability studies; this present research has some relation to usability studies and could easily be used in combination with other methods to bolster the present findings. As a relatively new feature, there has been no research to date specifically related to the Newspaper Search Interface.

Methods

Targeted recruitment email messages were sent to individuals responsible for Primo administration on each campus. A basic workflow and steps for retrieving the usage data of interest was created and tested by the researchers for clarity and ease of use. After initial testing, each campus representative was emailed *Data Extraction Instructions*, in the form of a Google document. (The *Data Extraction Instructions* document is available with the other underlying open data online.) The instructions included screenshots detailing how to navigate each administrative usage statistics retrieval system as well as providing data transfer and file naming conventions. While specific implementation dates for the Newspaper Search interface varied for each intervention campus, for simplicity, comparability, and generality of the instructions, a set date range was selected which included a timeframe of pre and post intervention dates that were applicable to all sample campuses. In one case, a campus was unable to extract newspaper usage data from Primo analytics, but the Director of the Unified Library Management System, extracted and submitted the needed data for the affected campus.

[INSERT TABLE 1 CAMPUS SIZE COMPARISON]

The Natural Experiment Framework

This study used a natural experiment framework to analyze the effects of implementing the Primo Newspapers Search on news discovery activity, as measured by Primo Analytics and COUNTER R4 Database Report 1 result clicks and record views of specific news collections held across all libraries in the sample. In understanding a natural experiment framework, it must be stressed that "natural experiments are neither natural nor experiments." (Dunning, 2012) Natural experiments are technically social science observational studies. However, when a natural experiment is studied with a robust research design it offers a key advantage over other observational study designs: the potential inference of a causal relationship. In this particular natural experiment, causality is examined by application of the Neyman-Rubin Causal Model (RCM).

The Neyman-Rubin Causal Model is a well-established analytical framework used to compare treatments (i.e. interventions) in randomized experiments (Rubin & Zell, 2018). Causal effects are defined according to a potential outcomes framework, with the effect being the outcome that would have been observed had any unit in a sample which was assigned to the control group been assigned to the intervention group, or vice versa. Since any sample unit can only have fallen into one group or the other in any particular experiment, the problem facing causal inference which the RCM addresses is a problem of missing data (Dunning, 2012; Rubin & Zell, 2018). Randomization limits the potential of confounding variables to affect assignment to the intervention or control group. With randomization, the RCM can be applied to the group averages to determine the average causal effect: the observed difference in averages (statistical means) between the study groups.

How then do analysts apply the Neyman-Rubin Causal Model to this observational study? The answer lies in a demonstration that allocation of the libraries in the sample was as-good-as or "as if" random. In this sample, the intervention group was not randomly allocated to receive treatment; they self-selected into the group by activation of the Newspapers Search. The randomly allocated control group assumed by the RCM was composed of the five libraries at campuses with full-time equivalent student population closest to a library in the intervention group, it is therefore referred to as the pseudo-control group. Given the necessary role in assuming "as if" random allocation between intervention and pseudo-control groups, detailed qualitative and quantitative analysis was done to justify the assumption.

Establishing "as if" Randomization

Qualitative Checks

The role of qualitative evidence in establishing the plausibility of "as if" randomness is central. - What grounds any hypothesis about what happened during the natural experiment and the casual processes at work is qualitative information. Such qualitative information comes in two forms which explain the data-generating process and ground a causal hypothesis in reality. First, the 'boots on the ground' insight from investigators and the data contributors that provide contextual information about why the intervention and pseudo-control groups ended up as such. In other words, why some campuses chose to activate the Newspapers Search and why others declined. Second, analysts must judge the intervention and pseudo-control groups theoretically according to the information, incentives, and the capacities of the units under examination (Dunning, 2012).

Information

First, investigation must be done to see if the units under study had information about their eventual assignment to the intervention or pseudo-control groups. Did the libraries implementing the Newspaper Search know they were being exposed to the intervention? Yes; they knowingly "exposed" themselves. Did their end users know this? At least the ones using the new Newspaper Search feature did. The subsequent question then becomes, did the libraries analyzed in this study know, during the data collection period, that they would later fall into either the intervention or pseudo-control groups? Did their end users know this? Only a very select few individuals at each campus library knew that a study of the Newspapers Search would be undertaken. Importantly, this knowledge was revealed to them after those in the intervention group decided of their own volition to activate the feature. Study plans informing campuses of the natural experiment design were communicated to campuses in the middle of the fall 2019 semester, at a time when it would be highly unlikely for them to change their Primo configuration, thus ensuring that for the majority of the data collection period, the units under study did not even have information that might lead them to affect the study design or data-generating process.

Incentives

Next, investigation must be done to see whether the studied units had any incentives to self-select into the intervention or pseudo-control groups. Did the libraries have incentives to activate the Newspapers Search? Yes; if they believed the marketing from Ex Libris that the Newspapers Search would either improve or simplify access, and they wanted such an experience for their users. Did these libraries have incentive to affect the study design or the data-generating process? No; as noted above, this study was conceived of in late 2019. By that time libraries had already sorted themselves into the intervention group and analysts paired them with pseudo-controls after

the fact. Would end users have any incentive to use the Primo instance of another CSU library - at an institution they were not enrolled in - in order to better discover newspapers? The Newspapers Search is a quick way to access news, but all of the campuses studied had access to the ProQuest Global Newsstream database which also provided quick access. End users were, as noted above, not aware of the existence of the study, nor were they necessarily aware that other campuses offered an easy way to discover newspapers within Primo. It is possible but unlikely that end users had any incentive to affect the data-generating process.

Capacities

Finally, investigation must be done to see if the studied units had the capacity to self-select into either the intervention or pseudo-control groups. Did libraries have the capacity to enter either group? In a sense, yes; any library could have activated the Newspapers Search at any point after it was available in the production release of Primo. However, there are reasons why an affirmative answer to this question does not affect the logic of this natural experiment. Mainly, the fact that the present study was retroactive. If any of the campuses in the pseudo-control group would have activated the Newspapers Search, they would have fallen in the intervention group and then a different CSU library with a similar FTE figure would have been selected for the pseudo-control group. The sample size of both groups would then have increased, rather than study design being compromised. Would end users have the capacity to use the Primo instance of another CSU library to discover and access newspapers? In reply to the 'discovery' question, the answer was yes; but in reply to the 'access' question, the answer was no. Each library's authentication system would have prevented students/faculty/staff at another institution from having the capacity to arrive at the full text of the newspaper articles. Analysts stipulated that study units in either of the groups did not have the capacity to affect the data-generating process for the COUNTER 4 data. As for the

Primo Analytics data and web analytics; while the theoretical capacity was possible, it seemed unlikely.

In this study, the qualitative checks clearly support the plausibility of "as if" random allocation. Why did the intervention campuses activate the new feature? They were either curious about it or believed it would offer a superior experience to the end users. Why was the pseudo-control group composed of the five particular campuses out of the larger group of CSU Libraries that did not activate the new feature? They were the closest in full-time equivalent student population to one of the intervention campuses and formed a logical pair. When considering the information available to, incentives of, and change capacity of the libraries under study (and their end users), analysts find no compelling reasons to reject the assumption of "as if" random allocation. Though the qualitative picture is compelling, the assumption of randomness is such a strong and weighty one that qualitative analysis alone cannot justify it; quantitative checks of the intervention and control groups are required as well.

Quantitative Checks

Carnegie Classifications

The primary quantitative checks on the validity of "as if" randomness come via an analysis of the Carnegie Classifications of Institutions of Higher Education data for the ten sample campuses. Carnegie Classification is the leading framework for describing and quantifying the diversity of higher education institutions in the United States. It makes use of the National Center for Education Statistics Integrated Postsecondary Education Data System (IPEDS) as well as data collected by The College Board (Indiana University Center for Postsecondary Research, n.d.). As such, it is the

most robust source of data and metrics that can be used to statistically compare higher education institutions. Using the most recent Carnegie Classification data, analysts compared the ten campuses in the sample along all available measurements. In total, intervention and control campus groups were compared along 68 quantitative and qualitative (which were coded to allow statistical analysis) variables. Tests of equality of means and statistical significance were performed for all variables with a two-sample Student's *t*-test; the independent categorical variable used to group the campuses being a dichotomous intervention or control value. Appendix Tables A, B, and C display the results; Appendix A shows general institutional characteristics, Appendix B shows institutional enrollment characteristics, and Appendix C shows institutional student characteristics. The *t*-statistic values were uniformly low and no statistically significant differences between the intervention and pseudo-control campuses were found indicating that "as if" random allocation of institutions to those groups can be plausibly assumed. Results of the quantitative check analysis for all variables in the Carnegie Classifications dataset are openly available online.

Database A-Z Listing Pageviews

An additional quantitative check on the validity of hypothesis and data comes in the form of web traffic analytics for each library's database A-Z listing page. All ten libraries in the sample maintain a webpage listing all of their electronic databases in alphabetical order or grouped by primary subject. As traffic analytics show, these webpages are well used and a considerable if not the primary point of access for users into each library's databases. To verify whether any potential increase in COUNTER 4 data was caused by the change of implementing the Newspapers Search and not due to a spike in traffic from the database A-Z list, page view count data were collected from each of the ten campuses for the same time periods as the COUNTER 4 and Primo Analytics data. Trendline data and difference-of-means calculations for each intervention and pseudo-control

campus pair were compared. Findings and implications of these comparisons are reported below in the Results section.

Comparison of Means & Significance Testing

After the "as if" random allocation of campus assignment to control or intervention groups has been successfully demonstrated, the quantitative analysis of a natural experiment can be simple. Dunning has explained that the most straightforward and compelling evidence of causal effect(s) is a difference-of-means (or difference-of-percentages) analysis combined with an applicable test of statistical significance and requisite confidence interval and standard error calculations (Dunning, 2012).[1] Importantly, the difference-of-means is calculated as the difference between the averages of the control and intervention groups, not between the intervention/pseudo-control campus pairs. The difference between the intervention group average and the control group average is an estimate of the average causal effect. Group averages are compared because individual campus potential outcomes under both treatment and control are unobservable since only one group assignment is possible. (A library cannot both have the Newspapers Search active continuously and not have it active continuously for the duration of the study period.) Similarly unobservable is the true average causal effect, which is formally defined as the difference in outcomes if every unit studied were assigned to intervention minus the outcome if every unit were assigned to control.

---

[1] Donald Rubin, of the RCM which bears his name, refers to this type of statistical approach as "Neymanism" (Rubin & Zell, 2018).

Fortunately, "as if" random allocation allows a simple and credible *estimation* of the average causal effect through reliance on the logic of random sampling and the statistical principle that the mean of a random sample is an unbiased estimator for the mean of the larger population (Dunning, 2012; Knapp, 2008). This estimation is obtained by the same method described above to obtain the average causal effect and the statistical standard error estimate is attached to the difference-of-means. The standard error being the standard deviation of the statistic's distribution which measures the random variation around a parameter, in this case the actual average causal effect, which it estimates. The size of the standard error is partially dependent upon the sample size; the larger the sample size, the smaller the standard error and the more closely the reported statistic will cluster near the actual parameter (Little, 2004). An estimate of the average causal effect and accompanying standard error are not sufficient to demonstrate a causal relationship, statistical significance testing must be performed as well. Data collection for this study spanned 26 months for each campus, in theory, which would have yielded a decent sample size. However, Ex Libris made the Newspapers Search available prior to their introduction of tracking capabilities for search activity in Primo Analytics. This resulted in a reduced sample size and missing data for Primo Analytics calculations; these factors are addressed in more detail in the Limitations section. The fact that most campuses implemented the Newspapers Search at slightly different times, combined with the missing data problem, left the analysts with unequal intervention and control sample sizes (with the samples here being monthly search activity totals) of the two variables, COUNTER 4 and Primo Analytics. This fact of unequal sample sizes violated a necessary assumption for the use of the traditional Student's *t*-test to compare pre- and post- treatment differences. Welch's unequal variances *t*-test, which does not require equal sample sizes, was thus used to test statistical

significance (Bacher, 2004).[2] Following standard procedures for the analysis of "as if" random natural experiments, the intervention campus group which turned on the Newspapers Search is compared with their pseudo-control pair campus group which did not. Specifically, analysts calculate the difference-of-means for COUNTER 4 and Primo Analytics data from the 10 campuses along with the statistical significance and standard error of the average causal effects; results are reported below. A brief discussion of the patterns observed via comparison of the web analytics follows.

Results

Average Causal Effect: COUNTER 4

[INSERT TABLE 2 COUNTER 4 INTERVENTION COMPARISON]

[INSERT TABLE 3 COUNTER 4 CONTROL COMPARISON]

COUNTER 4 Results Clicks increased after implementing the Newspapers Search at all five intervention campuses. In the pseudo-control group however, Results Clicks increased at three of the five campuses. The relative increase in Results Clicks among the pseudo-control group was larger than the increase among the intervention campuses leading to negative causal effect of 367.16 Result Clicks per month (p=0.02). With 99% confidence, the causal effect was between 420 and 313 fewer COUNTER 4 Result Clicks per month, based on 61 monthly samples.

---

[2] Welch's *t*-test also does not require equal population variances. It is therefore more robust and a more conservative test of the hypothesis that two normally distributed populations have equal means.

COUNTER 4 Record Views increased after implementing the Newspapers Search at three of the five intervention campuses. In the pseudo-control group, Record Views increased at four of the five campuses. The relative increase in Record Views among the pseudo-control group was larger than the increase among the intervention campuses leading to a negative causal effect estimate of 119.66 Record Views per month (p=0.14) which failed to achieve statistical significance. The 99% confidence interval for this finding was between 146 and 93 fewer COUNTER 4 Record Views per month, based on 61 monthly samples; again this result was not statistically significant. Results of the difference of means calculations, their statistical significance, and the standard error and confidence interval for COUNTER 4 Result Clicks are presented in Table 6 Average Causal Effect Estimates.

Average Causal Effect: Primo Analytics

[INSERT TABLE 4 PRIMO ANALYTICS INTERVENTION COMPARISON]

[INSERT TABLE 5 PRIMO ANALYTICS CONTROL COMPARISON]

Usage metrics from Primo Analytics reveal a clear pattern. Among the pseudo control group, newspaper facet usage increased at three of the five campuses and only by small amounts in each case. Because implementing the Newspapers Search scope in Primo removes newspaper content from other scope results, facet usage could not be used to track activity in the intervention group; newspaper scope search activity was used instead. At all five of the five intervention libraries, newspaper scope search activity increased post-treatment compared to their pre-treatment facet usage; the positive causal effect was 195.71 newspaper searches per month (p=0.01). With 99% confidence, the causal effect is between 160 and 231 additional Primo Newspaper Search searches per month, based on 40 monthly samples.

Results of the difference of means calculations, their statistical significance, and the standard error and confidence interval for Primo Analytics newspaper activity are presented in Table 6 Average Causal Effect Estimates.

[INSERT TABLE 6 AVERAGE CAUSAL EFFECT ESTIMATES]

Database Listing Web Analytics

Results from analysis of web traffic to the database A-Z list page at each library supported the hypothesis that implementing the Newspapers Search caused the increase in Primo Analytics usage statistics reported above. To recapitulate the theory behind collecting and reporting database listing page views: the purpose they served was an additional check on the posited causal relationship. If it was observed that web traffic to the database listing page increased at campuses that had an increase in COUNTER 4 usage, that would cast some doubt over the theory that an observed change in COUNTER 4 data was caused by implementation of the Newspapers Search. Obviously, web traffic data to a page listing all of a library's databases is a very crude measure, something elaborated on below. Nevertheless, average web traffic to the databases listing page at each of the intervention libraries decreased after the intervention. A Welch's *t*-test was conducted to determine if there was a statistically significant difference between the pseudo-control and intervention campuses in their database page traffic. Results of a pre-intervention comparison found a pre-intervention difference of 3,444 views per month (p=0.02) between the two groups and a post-intervention difference of -631 views per month which was not statistically significant (p=0.76). This empirical result demonstrates that the observed change in COUNTER 4 Record Clicks data

among intervention campuses was not caused by a shift in traffic to Global Newsstream and Ethnic Newswatch via their database listing page. Similarly, the small statistically insignificant increases in COUNTER 4 Record Views observed at four of the five intervention campuses cannot be attributed to a traffic shift involving the database listing page, traffic actually declined at four of the five intervention campuses; results are reported in Table 7 Database A-Z List Traffic Intervention Comparison. Note that a causal relationship was not posited here: while changing the Primo interface for better newspaper discovery could theoretically cause fewer visits to a library's database listing page, such a result would be unlikely because most students and faculty are looking for sources other than news. Similarly, it was observed that average web traffic to the databases listing page at four of the pseudo-control libraries decreased after the intervention at their paired institution. This finding, in conjunction with the COUNTER 4 data from the pseudo-control group lends credence to the theory that web traffic patterns did not vary wildly and thus that any change in newspaper usage came either as a result of changes to Primo or some other unobserved factor; results are reported in Table 8 Database A-Z List Traffic Control Comparison. Given that web traffic to the database listing pages decreased for eight of the ten libraries over the study period, the most reasonable inference is that the decrease was part of a secular trend or simply an artifact of when the study period fell in relation to the academic calendar.[3]

[INSERT TABLE 7 DATABASE A-Z LIST TRAFFIC INTERVENTION COMPARISON]

[INSERT TABLE 8 DATABASE A-Z LIST TRAFFIC CONTROL COMPARISON]

---

[3] Recall that data collection halted in February 2020 and therefore was a snapshot of 'normal' activity prior to the global outbreak of the COVID-19 disease.

Discussion

The authors of this study feel that there is a wide array of potential discussion points prompted by the study results, but have chosen to focus on the following four main issues: Study Limitations, Initial Assumptions and Data Scope, Issues not Explored, and Post Study Changes to Primo.

Study Limitations

While this study is robust, there are a few limitations that should be addressed including concerns surrounding the validity of the data. One of the tenants that makes this research and method unique is the use of quantitative usage data not qualitative data to draw conclusions about user searching behavior; but this leads to reliance on the accuracy of this data to correctly portray user behavior. Do analysts have enough data to make evidence-based decisions that reflect actual real-world scenarios?

Primo Analytics

There is concern within the Ex Libris customer base on the unreliability of Primo Analytics and while this product vendor continues to improve these systems, it is still under development and several bugs are known. Astute users of Primo Analytics know that the product has a number of limitations including multiple anecdotal reports of obtaining different metric counts when compared with Google Analytics and the comingling of Sandbox and Production instance data (Erhardt & McMunn, 2019). There are also questions about the reliability of the data itself; unexplained and uncorroborated spikes in PA data have been observed and as of 2019 there were dozens of support cases open with Ex Libris about inconsistent or missing Primo Analytics data

(Heller & Martin, 2019). One foundational limitation is that analytics tracking for usage of the Newspapers Search only became available as of the May 2019 production release of Primo (Natan & Yehuda, 2019). This was almost an entire year after the Newspapers Search was available for libraries to activate in their Primo production instances. This obviously impacted the sample data and the data-generating process for Primo Analytics itself since some of the intervention groups activated the Newspapers Search prior to May 2019. Thus analysts were faced with some missing data for certain months and a reduced sample size (measured in number of monthly totals of 'search' activity) which impacted the statistical standard error calculations and confidence interval. Nevertheless, analysts still found a statistically significant result ($p = 0.01$) so they do not believe this limitation affected the hypothesis that implementing the Newspapers Search will result in increased Primo search activity for news content.

Pageview Data

Actual 'clickthrough' data for the three ProQuest newspaper databases examined would have been a strong measurement. However, not every library in the sample could produce such 'clickthrough' data from their databases listing webpage at the micro level of an individual database. Analysts therefore were forced to rely on the only metric that could be collected from each member in the sample that was standardized so that reliable comparisons between institutions over time could be made: page views. While the researchers of this study would have liked to present an even richer picture of the way patrons discover and interact with newspaper content in the discovery layer, the detailed metrics which consumers of usability studies might be familiar with are simply not possible to obtain without the use of additional third-party tracking software. Since not all libraries in the sample were using such tracking software such as Google Analytics, this data was excluded

from analysis. Because this study only looked at the effect of implementing a separate webpage and query point for newspaper content, the results should not be extrapolated to the general question of the effect of segmenting content within a discovery layer. The 42% decrease in newspaper Result Clicks and 66% increase in search activity for newspapers observed here would not necessarily hold for other types of content.

COUNTER 4 Data

There are several limitations as well as validity concerns surrounding the collected usage data. One specific limitation of interest is the documented inherent unreliability of vendor provided COUNTER R4 data. COUNTER R4 measures have been extensively criticized by librarians and data managers for inflation due to duplication of the database structure as well as ambiguity of formats leading to current standards improvements with a new, evolved COUNTER R5 standard. This study could be improved with a longer time span and consistent data measures to analyze. With the introduction of more precise COUNTER R5 measures, there is a promise of a more accurate representation of full text downloads. While the inaccuracy of R4 data is well known, inclusion of COUNTER data was an important instrument to observe user behavior outside of Primo and better understand patterns of use, creating a full picture of how and when users accessed digital newspaper content. The superior COUNTER R5 Total_Item_Investigation measure and reports were not adopted by the ProQuest vendor with enough lead time to the intervention to provide a full view of activity. An initial comparison of R5 to R4 data did suggest that there were possibly more effective measures used and eventually the R4 data proved to be highly problematic as ProQuest divides the subset of the Global Newstream database into individual parts leading to extensive data cleanup to reveal a single monthly total.

Framework

Lastly, readers should note the Achilles' heel applicable to all natural experiments: the possibility that one or more unobserved confounding variables affected allocation to intervention or pseudo-control groups (Dunning, 2012). Such statistical confounding may have affected the difference in average outcomes between the groups on the variables analyzed and concomitant average causal effect measurements. While analysts consider the presence of unobserved confounders highly unlikely given the qualitative and quantitative checks performed to establish balance between the two groups and justify the assumption of "as if" randomness, it simply is an ever-present factor to be considered when using a natural experiment framework. Future research on this topic using actual randomization may find different measures of the causal effects.

Initial Assumptions and Data Scope

The first assumption made by researchers is that there is something of interest to learn from doing the research, and that the methods chosen may be able to provide insight on the research question at hand. From there researchers may have an intuitive guess as to what the data will show once collected and analyzed. Of note for this study is the fact that the three principal researchers had several differing initial assumptions about the outcome of their research, and that they were all wrong. One researcher expected the data to show that implementing the Newspaper Search would highlight the content, making it more visible and potentially increase usage. Two researchers expected the Newspaper Search implementation to decrease usage because it shifts newspaper discovery away from the main search results. This research showed that there was no increase in newspaper usage that could be tied to the implementation of the newspaper search, rather the statistically significant change that occurred with implementation was the locus of discovery.

Campuses that implemented the newspaper search had the bulk of their newspaper discovery happen in OneSearch, whereas control campuses had more discovery happen within each newspaper database, as measured by COUNTER 4 Result Clicks and Record Views, and Primo Analytics.

Had any single campus in this study examined their COUNTER 4 and Primo Analytics data, they would have seen an increase in newspaper usage over the study period, potentially leading implementation campuses to assume that implementation led to the increase in usage. However, through the natural experiment lens, there was no statistical increase in usage between the intervention and control groups. With a wider data view, it became clear that usage was not the metric that changed, but discovery location. A narrow data view can distort researchers' understanding of their research results.

Consortia are excellent laboratories for recruiting LMS natural experiment participants, as members have easy access to comparison data, and may be more aware of system implementation decisions made by consortia members. However, researchers can match any group of control and intervention institutions on FTE enrollment, Carnegie demographic data, electronic resource holdings, and implementation or not of a new LMS feature to identify and recruit participants for a natural experiment.

Issues not Explored in this Research

Qualitative Methods

This study could be improved as a mixed methods study that also considered users preference through surveys, focus groups, interviews, and/or direct observation. Qualitative research methods could further explore whether students understand the nature and format of different types of sources, such as the difference between a scholarly research journal or popular newspaper publication. Students' comprehension of digital literacy concepts has deep implications for how users discover and interact with digital content. Do students understand the benefits and disadvantages of a dedicated newspaper search interface? Are library decision makers making unfounded assumptions about searcher behavior and preference, such as undergraduates seeking newspapers as a separate entity from scholarly publications? While current undergraduates are perceived as "digital natives" comfortable with technology, it is impossible to have a discussion on digital newspaper use without also acknowledging the current climate surrounding "fake news" and the fear of inaccurate or misleading information, making library resource marketing another unexplored factor in newspaper usage.

Library Instruction

Not considered in this study, but highly relevant to an examination of user search behavior is the consideration of how students are taught to use discovery systems in library instruction. In theory students would use systems the way librarians teach and train them, but users come to libraries having prior experience with Google searching, and librarians must adapt to that prior experience

in order to compete in digital resource discovery. Out of scope to this work, but of interest is the influence of Google on the evolution of discovery layers and users searching expectations.

Post Study Changes to Primo

In this study, each consortia member campus made an independent decision to implement the newspaper search or not based on the information available about the newspaper search at the time, although CSU Bakersfield did turn on the search at the recommendation of Ex Libris support to address an indexing mismatch issue. Changes made to Primo after the conclusion of this study will likely make campuses that did not implement the newspaper search reconsider their decision. On June 15, 2020 the CSU went live with a new Primo function called the Central Discovery Index (CDI) , a replacement for the Primo Central Index (PCI). Both LMS elements are enhanced metadata indexes shared by all Primo instances. PCI and CDI data provide article and book chapter level indexing for many titles that otherwise wouldn't have that level of indexing available. Most differences between the PCI and the CDI are outside of the scope of this article, but one difference is very relevant to this discussion. Using the PCI, enhanced metadata for newspaper articles was treated and served no differently than journal article or ebook metadata. If the electronic collection package was available and turned on for PCI functionality, the information would show up in OneSearch within the main results set, and within the newspaper search. With the CDI, newspaper enhanced metadata will only show up in systems where the newspaper search is activated, it will not show up in the main search results. This research shows that turning on the newspaper search does no harm, in terms of cost-per-use or return on investment, as it does not lessen overall newspaper discovery. Campuses that wish to make CDI metadata for electronic newspaper collections available should turn on the newspaper search.

Conclusion

The ultimate goal of this study was to determine if separating newspaper content into a dedicated search interface improved the user discovery experience. The conclusion is that it shifted the majority discovery experience from within individual databases to occurring within the LMS, but did not impact the amount of discovery happening. Other actions, such as information literacy and library marketing efforts, are needed to seriously impact the total amount of newspaper discovery.

This research demonstrates a theoretical framework to approach evaluating commercial LMS release updates and optional features. Analysts applied the Neyman-Rubin causal model using a natural experiment to Ex Libris' Primo specifically, but the framework is versatile enough that it could be applied to the evaluation of new features in other catalogs or discovery layers. Web analytics from multiple sources were used to identify data/measures based on actual user searching behavior to determine the success or failure of alterations to the discovery layer and overall impact on the user's experience. This research provides a baseline to begin targeting library instruction and marketing efforts. Librarians must consider where to invest limited budgets; understanding how users interact with the LMS and electronic resources can lead to more informed decision making and ideally better user outcomes.

Call to Action

This research is valuable as a model for examining modern discovery systems. The natural experiment framework could be applied in library digital environments iteratively as products continue to evolve, relevant to evaluating digital discovery, usability studies, as well as collection development strategies. This study focused exclusively on understanding the effect implementing the newspaper search interface had on collection use. Other libraries can adopt a similar model to design their own natural experiments as an alternative to A/B testing in discovery systems.

Acknowledgements

Data Availability Statement

The data that support the findings of this study are openly available in figshare at http://doi.org/10.6084/m9.figshare.12702140.

Works Cited

Atanassova, R. (2014, August 20). Improving the discovery of European historic newspapers. *IFLA WLIC 2014*. Libraries, Citizens, Societies: Confluence for Knowledge, Lyon, France. http://library.ifla.org/id/eprint/1038

Bacher, J. (2004). Welch Test. In M. Lewis-Beck, A. Bryman, & T. Futing Liao (Eds.), *The SAGE Encyclopedia of Social Science Research Methods*. Sage Publications, Inc. https://doi.org/10.4135/9781412950589.n1085

Daramola, C. F. (2016). Perception and Utilization of Electronic Resources by Undergraduate Students: The Case of the Federal University of Technology Library, Akure. *American Journal of Educational Research*, *4*(5), 366–370. https://doi.org/10.12691/education-4-5-1

Dunning, T. (2012). *Natural experiments in the social sciences: a design-based approach* (5th printing). Cambridge University Press.

Erhardt, A., & McMunn, W. (2019, May 2). *Primo Analytics: A Primer*. ELUNA 2019 Annual Meeting, Atlanta, GA. http://documents.el-una.org/1894/

Ex Libris. (2018a, April 22). *Introduction and Frequently Asked Questions for Newspaper Search* [Product Documentation]. Ex Libris Knowledge Center. https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/Primo/New_Primo_User_Interface/011Frequently_Asked_Questions_for_Newspapers_Search

Ex Libris. (2018b, April 23). *Configuring the Newspaper Search Interface* [Product Documentation]. Ex Libris Knowledge Center. https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/Primo/New_Primo_User_Interface/Configuring_the_Newspaper_Search_Interface

Gooding, P. (2016). Exploring the information behaviour of users of Welsh Newspapers Online through web log analysis. *Journal of Documentation*, *72*(2), 232–246. https://doi.org/10.1108/JD-10-2014-0149

Heller, M., & Martin, C. (2019, May 3). *Making Practical Decisions with Primo Analytics*. ELUNA 2019

Annual Meeting, Atlanta, GA. http://documents.el-una.org/1876/

Indiana University Center for Postsecondary Research. (n.d.). *The Carnegie Classification of Institutions*

of Higher Education, 2018 Update Public File [Data Set]. Carnegie Classifications. Retrieved

May 8, 2020, from https://carnegieclassifications.iu.edu/index.php

Knapp, T. R. (2008). Unbiased Statistic. In P. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods*.

Sage Publications, Inc. https://doi.org/10.4135/9781412963947.n601

Little, J. S. (2004). Standard Error. In M. Lewis-Beck, A. Bryman, & T. Futing Liao (Eds.), *The SAGE*

Encyclopedia of Social Science Research Methods. Sage Publications, Inc.

https://doi.org/10.4135/9781412950589.n956

Meyer, E. T. (2018). *The Scholarly Impacts of Newspapers: The Guardian, Washington Post, Wall Street*

Journal, and New York Times (p. 31). Oxford Internet Institute.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3194632

Natan, N. (2018). *Primo May 2018 Highlights*. Ex Libris.

https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/Primo/Highlights/020Primo

_May_2018_Highlights

Natan, N., & Yehuda, C. (2019). *Primo Quarterly Update - May 2019*. Ex Libris.

https://knowledge.exlibrisgroup.com/Primo/Release_Notes/Primo/2019/001Primo_2019_Release

_Notes

Njeze, M. E. (2013). Use of Newspapers and Magazines in the Academic Pursuits of University Students:

Case Study of Covenant University. *Library Philosophy and Practice (e-Journal)*, 1-9.

https://digitalcommons.unl.edu/libphilprac/845/

Office of Public Affairs. 2020. "The California State University Fact Book 2020." The California State

University. https://www2.calstate.edu/csu-system/about-the-csu/facts-about-the-

csu/Documents/facts2020.pdf.

Pacy, A. (2014, August 14). Newspapers in the Digital Age: A Case Study in How Public Library Patrons

Read the News. *IFLA WLIC 2014*. Libraries, Citizens, Societies: Confluence for Knowledge,

Lyon, France. https://www.ifla.org/files/assets/newspapers/Geneva_2014/s6-pacy-en.pdf

Rubin, D. B., & Zell, E. R. (2018). Causal Inference. In B. B. Frey (Ed.), *The SAGE Encyclopedia of*

Educational Research, Measurement, and Evaluation. SAGE Publications, Inc.

https://doi.org/10.4135/9781506326139.n102

Wolfram, D., & Xie, I. (2009). A Longitudinal Study of Database Usage Within a General Audience

Digital Library. *Journal of Digital Information*, *10*(4), 1-19.

https://journals.tdl.org/jodi/index.php/jodi/article/view/304/505

Yoon, H. Y. (2016). User Acceptance of Mobile Library Applications in Academic Libraries: An

Application of the Technology Acceptance Model. *The Journal of Academic Librarianship*,

42(6), 687–693. https://doi.org/10.1016/j.acalib.2016.08.003

Appendix

[INSERT APPENDIX A GENERAL INSTITUTIONAL CHARACTERISTICS]

[INSERT APPENDIX B INSTITUTIONAL ENROLLMENT CHARACTERISTICS]

[INSERT APPENDIX C INSTITUTIONAL STUDENT CHARACTERISTICS]

Tables and Figures

Table 1 Campus Size comparison

Pseudo-Control	Full-time equivalent student population	Intervention	Full-time equivalent student population	Date of intervention*
CSU Monterey Bay	6,605	Sonoma State University	8,250	August 2019
CSU Stanislaus	9,217	CSU Bakersfield	9,920	April 2019
CSU East Bay	12,805	CSU San Marcos	12,389	August 2018
CSU San Bernardino	18,319	CPSU San Luis Obispo	20,698	September 2019
CSU Long Beach	32,673	San Diego State University	32,169	June 2018
* Denotes first month where Newspapers Search was turned on in production Primo, zero-ing out Newspaper facet usage.				

Source: Enrollment Dashboard: Institutional Research & Analyses, The California State University https://www2.calstate.edu/data-center/institutional-research-analyses/Pages/enrollment.aspx All data from Fall 2019.

Appendix A General Institutional Characteristics

Carnegie Variable	*t*-Statistic	Degrees of Freedom	Significance (two-tailed)
2018 Carnegie Basic Classification	-1.34	8	0.22
2018 Undergraduate Instructional Program Classification	0.97	8	0.36
2018 Graduate Instructional Program Classification	0.63	8	0.55
2018 Undergraduate Profile Classification	-0.14	8	0.89
2018 Size and Setting Classification	0.00	8	1.00
Degree of urbanization (Urban-centric locale)	0.37	8	0.72
Hispanic Serving Institution	0.00	8	1.00
Minority Serving Institution	0.00	8	1.00
Council of Public Liberal Arts Colleges Member	1.00	4	0.37†
Coalition of Urban and Metropolitan Universities Member	0.00	8	1.00
Bachelor's degree total	-0.13	8	0.90
Master's degrees conferred	-0.55	8	0.60
Doctoral degrees - research/scholarship	1.00	4	0.37†
Doctoral degrees - professional practice	0.00	8	1.00
Doctoral degrees - other	-1.73	8	0.12
Total degrees conferred	-0.21	8	0.84
Research/scholarship doctoral degrees in Arts & Science fields	1.00	4	0.37†
Research/scholarship doctoral degrees in professional/other fields	1.00	4	0.37†
Master's, doctoral-professional practice, and doctoral-other degrees conferred in the Arts & Sciences	0.00	8	1.00
Master's, doctoral-professional practice,	-0.93	8	0.38

and doctoral-other degrees conferred in the Professional fields			
Baccalaureate degrees conferred in the Arts & Sciences (first and second majors)	-0.30	8	0.77
Number of baccalaureate degrees conferred in professional fields	0.00	8	1.00
Institution confers research/scholarship doctoral degrees	1.00	4	0.37†
Total dormitory capacity (campus-owned, -operated, or -affiliated housing)	1.15	8	0.28
† Equal variances not assumed. All other reported p values passed Levene's Test for equality of variances and calculations assume equal variance.			

Appendix B Institutional Enrollment Characteristics

Carnegie Variable	*t*-Statistic	Degrees of Freedom	Significance (two-tailed)
2018 Enrollment Profile Classification	-0.58	8	0.58
Annual enrollment headcount, academic year 2016-17	-0.12	8	0.91
Fall 2017 Full-Time Equivalent enrollment (full-time plus one-third part-time)	0.09	8	0.93
Undergraduate total enrollment, fall 2017	0.15	8	0.88
Graduate total enrollment, fall 2017	-0.73	8	0.49
Undergraduate degree-seeking full-time enrollment	0.20	8	0.84
Undergraduate degree-seeking part-time enrollment	-0.29	8	0.78
Graduate full-time enrollment, fall 2017	-0.41	8	0.69
Graduate part-time enrollment, fall 2017	-1.17	8	0.28

Appendix C Institutional Student Characteristics

Carnegie Variable	*t*-Statistic	Degrees of Freedom	Significance (two-tailed)
Final ACT category (1=inclusive; 2=selective; 3=more selective)	1.18	8	0.27
Number of first-time entering students who submitted SAT score	0.88	8	0.41
Number of first-time entering students who submitted ACT score	1.42	5.162	0.21†
Combined number of students submitting SAT or ACT scores	1.14	8	0.29
SAT-Verbal 25th percentile score	0.83	8	0.43
SAT-Math 25th percentile score	0.82	8	0.44
Combined SAT-Math and SAT-Verbal 25th percentile scores	0.82	8	0.43
The ACT equivalent score for the combined 25th percentile SAT score	0.92	8	0.39
ACT Composite Score, 25 percentile	1.00	8	0.34
Derived 25th percentile ACT score, weighting both ACT and equated SAT scores by number submitted	0.96	8	0.37
† Equal variances not assumed. All other reported *p* values passed Levene's Test for equality of variances and calculations assume equal variance.			

Table 2 COUNTER 4 Intervention Comparison

Campus	COUNTER 4 Metric	Pre-Deployment Average (per month)	Post-Deployment Average (per month)
Sonoma	Result Clicks	115.84	121.86
	Record Views	69.68	45.14
Bakersfield	Result Clicks	221.81	253.2
	Record Views	73.88	83.5
San Marcos	Result Clicks	370.25	677.78
	Record Views	117.5	364.72
San Luis Obispo	Result Clicks	668.3	1277.33
	Record Views	149.65	165.83
San Diego	Result Clicks	389	405.7
	Record Views	187.5	174.15

Table 3 COUNTER 4 Control Comparison

Campus	COUNTER 4 Metric	Pre-Deployment Average (per month)	Post-Deployment Average (per month)
Monterey Bay	Result Clicks	78.37	192
	Record Views	41.37	47.14
Stanislaus	Result Clicks	361.56	253.5
	Record Views	106.13	105.7
East Bay	Result Clicks	638	693.06
	Record Views	162.75	239.83
San Bernardino	Result Clicks	386.2	404.33
	Record Views	116.8	123.5
Long Beach	Result Clicks	1235.17	1749
	Record Views	431	652.4

Table 4 Primo Analytics Intervention Comparison

Campus	Pre-Deployment Average (per month)	Post-Deployment Average (per month)
Sonoma	207.84	369
Bakersfield	169.81	182.56
San Marcos	149.38	483.56
San Luis Obispo	176.95	640.67
San Diego	252.33	802.89

Table 5 Primo Analytics Control Comparison

Campus	Pre-Deployment Average (per month)	Post-Deployment Average (per month)
Monterey Bay	121.63	127.57
Stanislaus	206.88	194.1
East Bay	110	89.33
San Bernardino	220.95	256.67
Long Beach	600.17	602.15

Table 6 Average Causal Effect Estimates

Variable (Monthly Totals)	Average Causal Effect Estimate	Significance (two-tailed)	Standard Error	99% Confidence Interval
COUNTER 4 Result Clicks	-367.16	0.02	162.07	±53.45
COUNTER 4 Record Views	-119.66	0.14	81.06	±26.73
Primo Analytics newspaper activity	195.71	0.01	86.56	±35.25

Table 7 Database A-Z List Traffic Intervention Comparison

Campus	Pre- Average (per month)	Post- Average (per month)
Sonoma	2546.32	2485.71
Bakersfield	2174.50	1904.90
San Marcos	16,989	22,494
San Luis Obispo	11681.85	11127
San Diego	10237.83	8582.4

Table 8 Database A-Z List Traffic Control Comparison

Campus	Pre- Average (per month)	Post- Average (per month)
Monterey Bay	4,694	2718.71
Stanislaus	4541.94	3538.90
East Bay	8403.25	6838.22
San Bernardino	20,287	25,942
Long Beach	19497.83	19253.3