



The Case of the Missing Books: Using New Digital Analytics Data to Answer Old Questions

*Tabatha Farney, Matthew Jabaily, and Rhonda Glazier**

Introduction

When a book or other item goes missing in a library's collection, a decision must be made to replace the item or use the money to purchase a new item. Most libraries have developed a procedure to replace missing items, but there is not a single, standardized approach to this issue. To make this decision, age of material or fit for the collection is important, but so is understanding user interest in those missing items. Measuring user interest in a missing item can help libraries understand how relevant that item is to the library's collection. However, existing data practices often prove insufficient for measuring user interest for these missing items. For example, circulation statistics can be used to determine if the item circulated in the past, but items can disappear either the first time they are checked out or before they get a chance to circulate at all. Another potential method to measure user interest involves counting the number of holds on missing items or requests via interlibrary loan, but while libraries are often tracking requests, they are not tracking the number of times users find missing items in their library catalog or shelves but do not place a request.

Looking beyond traditional library data points, this study adapts a digital analytics approach to provide a more comprehensive picture of user interest in missing items. Digital analytics data has traditionally been used to assess user experience with a website, but it can also be implemented on an online library catalog or discovery layer for collection development purposes. This study seeks to determine if the number of pageviews on a missing item record helps identify which missing items to replace and whether this additional data point helps libraries make a better informed decision about replacing lost materials.

Literature Review

There is no standard process to determine when and how missing or lost items are replaced in libraries. However, the literature does demonstrate some potential methods. Smith et al. provides descriptions of a workflow and database created to automate missing book replacements in response to a growing backlog of missing items.¹ They formed a missing items task force to address item replacement. Because the missing items were ones that patrons had recently used or requested, they considered them high-use materials and shortened the search time for missing books from one year to six months. The article describes the design and creation of a database for tracking missing items and facilitating replacement decisions by subject librarians. While the system did help selectors identify whether there were other copies of the item or if newer editions were available, it did not include information about previous use of the items. After the initial adop-

* *Tabatha Farney, Director of Web Services and Emerging Technologies, University of Colorado Colorado Springs, tfarney@uccs.edu. Matthew Jabaily, Electronic Resources and Serials Librarian, University of Colorado Colorado Springs, mjabaily@uccs.edu. Rhonda Glazier, Director of Collections Management, University of Colorado Colorado Springs, rglazier@uccs.edu*

tion of the database, librarians processed 5,141 items and ultimately decided to replace 2,547 (49%) of those items.

Marien and Mundt describe another workflow, implemented at the American University Library, for the replacement of missing books.² In order to address a large backlog of replacement decisions, they decided to adopt replacement criteria that considered the age, price, and circulation history of the book. Notably, they automatically repurchased items published less than five years ago and automatically rejected any titles that were published more than 15 years ago and had not circulated in the last 15 years. Other replacement decisions were forwarded to subject specialists. Only items that had been lost for over one year were eligible to be replaced, and the replacement process was done once a year.

Many libraries use Google Analytics, a popular digital analytics tool, to track use of their websites. Some libraries also use it to track use of their online catalog or discovery services. In an early study, Fang and Crawford used Google Analytics to gain insights into patrons' use of the catalog and measure its usability.³ Their results focused primarily on characteristics of their patrons and the types of searches they conducted. While other studies have focused on search behavior, to our knowledge there are no published studies that use Google Analytics for tracking individual book titles for collection development purposes.

Background

The traditional method to replace lost materials in the Kraemer Family Library (KFL) differs if the item has the status of lost or lost and paid. For lost status items, the circulation department creates a list of missing status items each month and searches the stacks for these items. If an item cannot be found after the third or fourth search, usually over a four-month period, the Director of Collection Management is notified the item is lost and should be considered for replacement. Lost and paid status is designated for an item when a library user has checked it out, reported it lost to the circulation department, and paid the replacement fee. When the circulation department reports lost and paid status item to the Director of Collection Management, she usually waits three to four months before deciding if the items should be replaced. There have been several instances when lost and paid items were returned to the library after being replaced, so the waiting period was implemented to give these titles a chance to be returned.

The entire process to replace lost or lost and paid items takes several months—many times six months or more—with the only exception being for “high demand” titles that are on reserve. Reserve items are expedited through the process and replaced immediately. The assumption is that if a missing copy of a reserve item resurfaces, having an additional copy will increase access to the title and not be a waste of limited funds. Overall, the library replaces an average of 264 items each year using this process.

Replacing lost items with this method is time consuming and lengthy. It also raises the question of how the lack of access to these lost items impacts library users. To answer this question, the Director of Web Services and Emerging Technologies created a custom report using pageview data collected using Google Analytics to determine how often users viewed an item record in the library catalog that had a lost or lost and paid status. Google Analytics was selected for this study because it was already tracking website use data for the library catalog.

KFL began using Google Analytics' custom dimension feature to collect and report the title, call number, status, and location of items viewed in the online catalog in July 2017.⁴ With this data, it was possible to create a report of the items viewed by users that were marked either lost or lost and paid. With this data KFL could monitor the frequency users viewed a lost item to understand if users proceeded to request the item through interlibrary loan or were turned away from the item. The data suggested that reliance on interlibrary loan to provide access to unavailable items was insufficient to meet user needs, as less than 10% of users who viewed an

unavailable item clicked the link to initiate an interlibrary loan request. Given this evidence of significant, unmet demand, KFL wanted a better, more efficient method of replacing missing items and saw potential in the Google Analytics' data to assist in this process.

Methodology

To determine the effectiveness of replacing lost items using unique pageview data, the Director of Collections Management and Director of Web Services and Emerging Technologies each generated a list of lost or lost and paid status items.

The Director of Collections Management used the traditional replacement method of using a list of lost items generated by the circulation department and reviewed each title to determine if it should be replaced. Criteria used to decide whether to replace an item included the number of checkouts the item had before going missing, the publication date of the item and the cost of the item. Items where there were other available copies or newer editions of the same title were not replaced.

The Director of Web Services and Emerging Technologies created a custom Google Analytics report that contained the item title, call number, status, pageviews and unique pageviews for any item record viewed with a lost status from July 2017 to August 2018. The unique pageviews metric was selected to measure user interest as that data point most closely aligns to an individual session. Any item that had at least one unique pageview and did not have a second copy (either print or electronic version) was recommended for replacement to determine if unique pageviews could be an indicator if a replaced item would circulate. Item date, replacement cost, and prior checkouts were not considered when compiling this list.

These two lists were then combined and sent to acquisitions for ordering in August 2018. Most items were received between August 2018 and October 2018 with a few outliers. Items were replaced with the same edition unless a newer edition was available. In those cases, the latest edition was ordered rather than replacing the exact edition that was missing. For the purpose of this study, notes were added to the records of the items replaced to denote the method, or methods, used to select them. At the end of December 2018, circulation information was exported for all of the replaced items to determine how many times each had been checked out. A comparison group of new items purchased over the same period was also generated, so circulation data could be compared between the replaced and new items.

Results

Reports from Google Analytics showed that there were unique pageviews for 180 individual items that had lost or lost and paid statuses. These items had a total of 450 unique pageviews. By comparison, available items in the catalog had over 50,000 unique pageviews during the same period.

A total of 213 items were replaced. Of these replaced items, 131 were uniquely identified using the traditional replacement method, 56 titles were uniquely identified using the unique page method, and 26 titles were on both lists. Of the 213 replaced items, 82 titles (38.5%) circulated during the study. These items received 140 checkouts with another 33 renewals.

Analyzing the results by selection method found the items selected with the traditional replacement method had a 41% circulation rate and the unique pageviews selected items had 39% circulation rate. Replaced items on both lists were the most likely to circulate with 54% of those items circulating during this study. The comparison group of newly purchased titles from the same period contained 2,123 items, of which 361 (17%) circulated.

Of the unique pageview selected items, 72% had one unique pageview during July 2017 to August 2018. Only six items (7%) received more than five unique pageviews during this period. The 52 replaced items identi-

FIGURE 1
Number of Items Replaced by Method

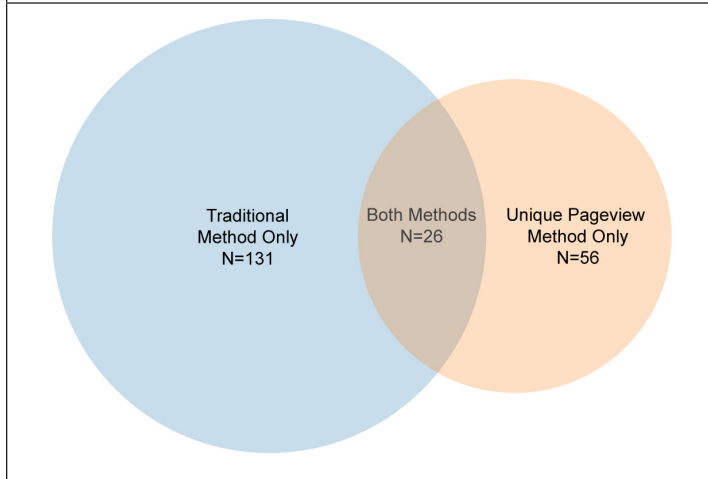
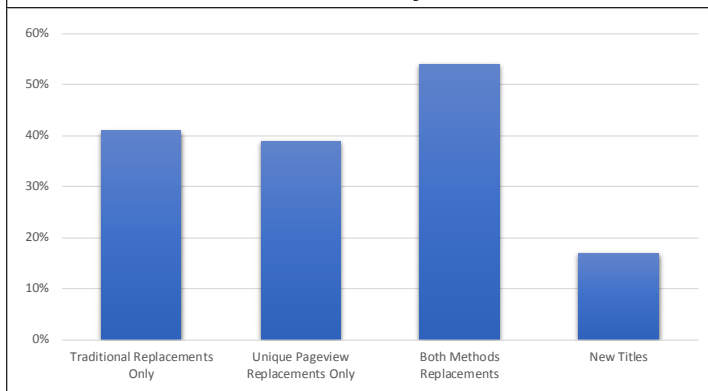


FIGURE 2
Circulation Rate by Method



fied with only one unique pageview had a 37% circulation rate and items with more than one unique pageview had a 43% circulation rate. The small subset of items that appeared to have high user interest with over five unique pages had a 33% circulation rate.

Findings

These results suggest that the addition of unique pageviews to identify lost items was worthwhile. Although the unique pageviews of lost items were a small percentage of overall unique pageviews, they represent a significant number of poor user experiences where users were not able to access the item they wanted. Although these items might have been acquired through another means, the data shows users did not use the link on the records to request an interlibrary loan. In light of this new information, the library has begun to prioritize replacing items more quickly or suppressing the item records from public view, as the clutter of unavailable items in the catalog can lead to further frustration for users.

Using unique pageview data this study increased the number of replaced items by over 25%, providing library users with access to

more items they are likely to use. Overall, the traditional replacement method and the unique pageview method selected items circulated at a rate significantly higher than that of the new items purchased during the same period. Items identified by both replacement methods were the highest performing, with over 50% of those items circulating during this study. This suggests that lost items identified using both methods can automatically be replaced rather than waiting several months before a decision is made.

While the two replacement methods had comparable circulation rates, the traditional replacement method found more lost items than the unique pageview replacement method. Many titles that were replaced using the traditional method circulated even though there was no prior evidence that they were viewed in the catalog. As a result, the unique pageview based method should be considered a supplement to the traditional method rather than a replacement.

A benchmark based on user interest using unique pageviews could not be established during this study. The original assumption was that lost items with higher unique pageviews would receive more check outs. Items that received one unique pageview had a 37% circulation rate, while items that received more than one unique pageview had a 43% circulation. The small subset of items that received over five unique pageviews only had a 33% circulation rate. However, items that were identified using both methods and had more than one unique pageview had a 63% circulation rate, suggesting KFL could automate the replacement process for lost items that meet these criteria.

The lack of an identifiable benchmark based on unique pageviews alone could be due to a notable limitation of the study: time. The replaced titles only had, at most, four months during the fall semester to circulate. Titles that took longer to enter the collection had less time to circulate, and it was not possible to accurately account for the difference in availability. Additionally, in order to identify a large number of potential replacement titles, the unique pageview data used was gathered over a 13 month period. The KFL's campus has three semesters (fall, spring, and summer) throughout the academic year and some courses are only taught one semester a year. As a result, missing items that received high pageviews over the spring or summer semesters may not have high user demand during the fall semester used for this study. Having a full year of data would be ideal, and the authors intend to reanalyze the data to determine if time period does impact user interest and usage of library items.

There are also limitations associated with using catalog pageviews as a measurement for user interest because it may underrepresent user interest. Google Analytics was configured to only record title and item information when a user views a record page, but in some cases, users may have discerned that the item they wanted was unavailable based on information from a list of results in the catalog or on a separate platform, such as the library's discovery layer. Second, users can opt to have Google Analytics not track their pageview data by disabling cookies or using an assortment of online advertisement blockers like Privacy Badger. KFL does not discourage users from using these options because the library believes users should be able to opt out if so desired. There is also the possibility of overestimating interest in lost items by collecting internal use data generated by library employees. To mitigate this problem, the authors of this study used a browser plugin to hide their activity from Google Analytics. They were also able to identify activity from at least one internal user, which was removed from the Google Analytics data.

Conclusion

Overall, the need to collect more data and expand the study is an important next step. While this study focused on lost and lost and paid items, broadening the study out to include items with a missing status in the catalog would be an effective next step. It will also be important to expand the timeframe used to study the replaced items to determine whether the initial interest in these items continues, or if the circulation drops off after the first circulation and becomes closer to the circulation rate for the overall collection.

Despite the limited amount of time and data, this study suggests using unique pageviews to measure user interest in an item has the potential to inform other collection development decisions. For example, this data could identify cases where additional copies of items are needed. Collection maintenance is an ongoing process in all libraries and tools that help identify issues that frustrate users are needed. Combining traditional and new methods of collection management will help all types of libraries better serve their users.

Notes

1. Cheri Smith, Anastasia Guimaraes, Mandy Havert, and Tatiana H. Prokrym, "Tutorial: Missing Items: Automating the Replacement Workflow Process," *Information Technology & Libraries* 28, no. 2 (2009): 93-99.
2. Stacey Marien, and Alayne Mundt, "Let's Get Technical—Nancy Drew and the Case of the Lost Books?" *Against the Grain* 8, no. 4 (2016): 88-89.
3. Wei Fang, and Marjorie E. Crawford, "Measuring Law Library Catalog Web Site Usability: A Web Analytic Approach," *Journal of Web Librarianship* 2, no. 2 (2008): 287-306.
4. Tabatha Farney, "Using Digital Analytics for Collection Development," in *Using Digital Analytics for Smart Assessment* (Chicago: ALA Editions, 2018), 70-72.

Bibliography

Fang, Wei, and Marjorie E. Crawford. "Measuring Law Library Catalog Web Site Usability: A Web Analytic Approach." *Journal of Web Librarianship* 2, no. 2 (2008): 287-306.

Farney, Tabatha. "Using Digital Analytics for Collection Development." In *Using Digital Analytics for Smart Assessment*, 63-80. Chicago: ALA Editions, 2018.

Marien, Stacey, and Alayne Mundt. "Let's Get Technical—Nancy Drew and the Case of the Lost Books?" *Against the Grain* 8, no. 4 (2016): 88-89.

Smith, Cheri, Anastasia Guimaraes, Mandy Havert, and Tatiana H. Prokrym. "Tutorial: Missing Items: Automating the Replacement Workflow Process." *Information Technology & Libraries* 28, no. 2 (2009): 93-99.