Library Catalog as a Tool for E-Book Discovery and Access in Patron-Driven Acquisition (PDA): A Case Study

Cristóbal Urbano
Universitat de Barcelona
Melcior de Palau, 140, 08014 Barcelona, Spain
urbano@ub.edu

Yin Zhang
School of Library and Information Science
Kent State University, Kent, OH 44242
yzhang4@kent.edu

Kay Downey
University Libraries
Kent State University, Kent, OH 44242
mdowney1@kent.edu

Tom Klingler
University Libraries
Kent State University, Kent, OH 44242
tk@kent.edu

ABSTRACT
In January 2012, Kent State University Libraries implemented a 6-month pilot project for a Patron Driven Acquisitions (PDA) e-book purchasing model that uses a combination of the acquisition services provided by the primary book jobber (YBP) and the access services provided by a well-known e-book distributor (ebrary). Using the book jobber mediated PDA model provided the library with selected discovery records that closely matched specifications of the library’s print approval plan. This poster presents a study that examines the role of the library catalog as a tool for e-book discovery and access in a new library collection acquisition model called Patron-Driven Acquisition (PDA).

Keywords
Patron Driven Acquisitions (PDA), e-books, library catalog.

INTRODUCTION
During January-June 2012, the Kent State University (KSU) library implemented a pilot initiative for a Patron Driven Acquisitions (PDA) e-book purchasing model, which uses a combination of the acquisition services provided by their primary book jobber, Yankee Book Peddler (YBP), and the access services provided by an e-book distributor (ebrary). PDA refers to the automated purchasing of e-books based on patron use. Through this model, the library provides access to a predetermined set of e-books to authorized KSU library users. Patrons discovering e-books via the library catalog have no clue whether a particular title is a PDA e-book available for purchase, or an item that has already been acquired. Once the patron’s usage of an e-book exceeds a certain threshold, the library automatically buys the e-book. In this way, PDA is a seamless selection process that augments traditional library approval purchases.

E-books at KSU can be found via a catalog search. A full bibliographic record display serves as the starting point for the triggering process. However, users can shortcut that “canonical” way by accessing the ebrary portal from links available in the list of results after a search in the catalog, as well as through the browsing and further searches performed directly in ebrary interface, once authenticated.

Despite all of the published literature—and vendors’ ads—that stress the key role of the library catalog in the PDA model for e-book acquisition, a literature review shows that very few log analyses have been undertaken to determine the actual usage of the catalog as a discovery tool to access the pool of e-books available in PDA programs. In order to gain a better understanding of how the catalog and library users interact in PDA, a comprehensive catalog web log analysis was conducted with catalog web logs during a KSU PDA pilot project that ran from January 1 to June 30 2012. This study aimed to examine two aspects of a library catalog in e-book PDA: (1) its function as a discovery tool for e-books purchased from user triggering actions, and (2) its function as a way to assess the overall usage of e-books in the PDA discovery pool (purchased or not).

PROBLEM STATEMENT
The discovery pool of PDA e-book records in the catalog is a key point in the PDA acquisition model for e-books, but up to now, few empirical works have studied how users interact with their catalog before they access the e-books at the provider platform. To improve the PDA model, manage the technical services workflow for e-books more...
effectively, and review the acquisition policies framing the picture of PDA purchased e-books in the overall collection evaluation, we need to know the actual use of the catalog and to contrast it with the general e-book usage reports that providers like ebrary supply. Such findings could provide information about new parameters to consider in the formula agreed upon with e-book providers, in order to tune the triggering process for PDA in future negotiations.

It is important to stress that commercial e-resource providers usually do not supply in their standard statistical reports any information about referral sources that bring library users to their webservers. This is the case with ebrary, meaning that the only way to approach how KSU users reach PDA e-books is the analysis of KSU library catalog web logs. KSU does not have any federated search tool that could compete with the catalog, nor are ebrary e-books indexed in search engines like Google and linked to the provider with a local link resolver. Therefore, the only options to discover PDA e-books available for KSU users are the library catalog and the browsing and searching options of the ebrary portal itself. In this work, we do not study non-cataloged sources of traffic to ebrary content, but only the level of e-book use linked to catalog searches and full bibliographic record displays.

LITERATURE REVIEW

Due to the acceleration of e-book acquisitions during the last few years, the literature devoted to PDA has experienced a great revitalization that can be traced through the tens of case study papers published on the issue during the last three years, as well as through some papers and books with a good general review approach to e-books (Blummer & Kenton, 2012; McLure & Hoseth, 2012). The Association of College and Research Libraries actually recognizes patron driven e-book acquisition as one of the “2012 Top Ten Trends in Academic Libraries” (ACRL Research Planning and Review Committee, 2012).

PDA is a trending topic in LIS literature, but few works have focused on user interaction with PDA records within library catalogs (or other discovery tools), despite the fact that it is an acquisition system that relies on the local catalog as the key discovery tool to access commercial platforms offering PDA titles. Although there has been a long history of studies on the use and users of OPACs since their introduction in libraries, with most of these using transaction logs analysis (TLA) methods (e.g., Borgman, Hirsh, & Hiller, 1996; Jansen, 2006), this approach has not been used to study the use of the library catalog as a way to access e-books, nor to study PDA programs.

We can also observe a rise of studies within the last few years on e-book use and users that show a mixed image of success and failure of e-book use in libraries (Slater, 2009). While some studies focus on the ways users discover and access e-books, few works have traced the users’ path from the OPAC to the commercial provider platform; an interesting and isolated example is the Colorado State University web survey applied through links embedded in the bibliographic records of a PDA set of e-books (McLure & Hoseth, 2012). In addition, a few libraries with some non-cataloged e-books in their collections have discovered, through the statistical reports from their commercial e-book providers, that cataloged titles were actually used more than those that were non-cataloged (Sprague & Hunter, 2008).

However, the statistical data about e-book use that libraries receive from commercial providers usually follow the COUNTER guidelines agreed upon as “de facto” standards for e-resource stats by libraries and the publishing industry (COUNTER, 2006, 2012), and do not offer information that could be obtained with a more detailed analysis of the row log data from the publishers’ webservers. It is therefore impossible to study the “referrer” log field that could help to determine whether the catalog or other library websites are the source of some visits to licensed content. It is also impossible to study the IP field that could help to map the uses from the library, both on- and off-campus. Few works have done this exploitation of raw log files, because a publisher or a consortial hosting involvement is necessary: for e-book content this is the domain of the now old OCLC Netlibrary study (Connaway & Snyder, 2005), and for journal content the domain of OhioLINK EJC (Huntington, Nicholas, Jamali, &Watkinson, 2006).

There is also a general agreement in works based on surveys and other direct methods (Blummer & Kenton, 2012) that the presence of e-books in the library catalog increases their use. For example, at the University of California (Chan, Poe, Poter, Quigley, & Wilson, 2011), 53% of users said that the library catalog was their main avenue to discover and access e-books. These studies provide a picture where the catalog is the main channel to access or to promote e-books, but the figures obtained are not very high and indicate that there is room for a greater level of involvement by the library in promoting e-book collections and the catalog as the vehicle of access.

Management of e-books in academic libraries is raising key challenges to library workflow, procedures, and policies (Vasileiou, Rowley, & Hartley, 2012). From the perspective of acquisition and technical services, PDA is a new, important challenge that should be studied in relation to the cataloging and workflow changes that are required to upload mass amounts of temporary records with uneven quality (De Fino & Lo, 2011). Promoting e-book use and performing PDA programs has returned online library catalogs to the spotlight. However, this shift has occurred as the debate about the future of the catalog is gaining momentum in the broader, fuzzy fields of web searching, metadata harvesting, and the umbrella term “discovery tools” (Chad, 2011). What is interesting to observe is that while PDA for e-books relies on the library catalog as a key piece for discovery, the role of the catalog as a discovery tool in that web scenario is still unclear (Kortekaas, 2012; Luther & Kelly, 2011).
METHODS

Data Sample
We traced the overall use of the catalog for a set of 20,062 e-books that have been available for the pilot test developed from January to June 2012: some titles uploaded later than January 1, and some titles discarded during the test, have not been considered for the study. The period of analysis for the usage of e-books reported by ebrary is delimited by that pilot period, plus the next six months of usage until December 31, to follow up on use in the next academic term. However, for the 518 e-books purchased through PDA during the pilot period, we checked logs in the exact date of triggering in order to know whether the triggering process is linked to an actual search in the catalog.

Activities Recorded in the Catalog Webserver Logs
All catalog search transactions during January 1, 2012 to December 31, 2012 were gathered to extract all e-book related searching and access activities.

Data Processing
We obtained the catalog record ID for the full list of 20,062 PDA e-books, in order to trace them to the KentLink log file. A batch program has been performed to extract all of the catalog log lines where a record ID of one of the ebrary e-books appears in the “Request” field of the log line. Every time a request of an ebrary book is made from the full bibliographic record in KentLink, the log file records two lines, which include the KentLink ID and the search query embedded. We considered a period of 30 minutes to be an indication of a time-out within the sessions of searches in KentLink logs. In order to define the unique clicks to an e-book we have considered this time-out as well as any changes for each KentLink bibliographic record ID in the log line that represents a different value for IP, Request, Referrer, Agent, or Port.

In order to establish when a search in the catalog matches with a trigger action on a given date, we filtered the logs by IP. Only IPs from the KSU range, which includes VPN connections outside the campus, can trigger the purchase, due to the IP recognition authentication that occurs when the user follows the link to the ebrary content. Users with external IPs cannot access the full text content in the ebrary KSU portal, nor can they trigger a purchase following the link from the catalog.

We worked with two ebrary sources: the Title Report for the monthly use of each book and the Trigger Report, which supplies for PDA purchased books such information as the first use of each title, the trigger date, and the type of activity performed that triggered the purchase. What is important about the Title Report is that it offers information on the use of non-purchased PDA e-books as well as those that are triggered. A weak point from the ebrary stats is that up to now, they do not offer any information on the number of users registered with a customizable profile in their portal, which could prove fundamental for the analysis of the usage originating from tablets and mobile devices.

We matched the Trigger Report with the logs to trace whether a full-orthodox search had been performed from a KSU IP, while taking into account the time zone disparity between the KentLink server (US Eastern Time Zone) and ebrary server (US Pacific Time Zone). Since the Trigger Report offers only the date and not the time of the triggering, when more than one full-orthodox search occurred in the trigger date, we considered all of the searches performed that day as potential clicks for triggering, and analyzed them in a separate set from the cases with an unique IP/search as a triggering agent.

FINDINGS

General catalog use and types of searches: Out of the 20,062 e-books in the PDA discovery pool, 3,254 occurrences were from regular catalog searches and e-book access links from the full display record (“full-orthodox”). These 3,254 access occurrences represent 2,369 different unique searches. Not surprisingly, keyword search was the most prevalent search type performed, with title search coming in second.

E-books purchased and used as reflected in the catalog logs: Out of 20,062 e-books in the discovery pool, only 1,131 (5.64%) have been consulted, and 518 (2.58%) have been triggered for purchases during the 6-month pilot period. It is interesting to note that 22.01% of the triggered books never have been accessed from a catalog full record compared to 12.29% for the general sample pool.

E-book use by number of clicks up to purchase triggering: The number of clicks before triggering has a strong relationship with the intensity of use during the observation year (six-month pilot period plus six months after pilot).

E-book use by search type at the triggering point: It is important to assess the level of use by search type, as we assume a more elaborated, or specific, search may better present the real information needs of users. From the books with triggering points clearly linked to a specific search in the catalog, we examined the average usage by search type of 361 total titles. The results show that author and title searches with a direct link to the triggering of an e-book led to more uses for those titles. Furthermore, as many keyword searches are title-like (with all of the words of the titles, but using the default search form of the catalog), keyword searches linked to the actual triggering of some books are related to a similar level of use.

DISCUSSION AND CONCLUSION
The two main objectives of this work were to examine whether the full use of the catalog was the main gateway to access PDA e-books at the vendor site, and whether the higher and more intense e-book use was linked to full catalog interactions at the triggering point. While results of
the study provide positive answers to these questions, they also leave some uncertainty. Indeed full catalog interaction matters in PDA because our findings show that more than 70% of the e-books were purchased after a trigger action linked with a catalog search and a full bibliographic record display and that the use of e-books were always slightly higher when a catalog search is involved.

The results of this study show that the library catalog plays an important role in e-book discovery and access. As library catalogs normally only show what libraries have purchased and what is actually in the library collection, the low usage and low triggering rate of e-books in the PDA discovery pool suggests a more targeted e-book PDA pool available in catalogs. The results of this study show that catalog search types (such as keyword and title) are closely related to e-book triggering, access, and use. The highest average number of user sessions is linked to author search and the highest for page views corresponds to title search. This study suggests follow-up user studies of the catalog interface design for more effective e-book discovery and access. Additionally, further analysis of keyword searches is also needed to explore effective ways for e-book discovery and access. For example, splitting the “title like” keyword searches from the rest of the general, more subject-oriented keyword searches would help review subject search as a viable option for e-book discovery and access.

ACKNOWLEDGEMENTS
Cristóbal Urbano received a grant for this research from the Spanish Ministerio de Educación, Cultura y Deporte, 2012-2013 Programa Nacional de Movilidad de Recursos Humanos del Plan Nacional de I-D+i.

REFERENCES