

Facilitating Content Discovery and the Value of the Publisher Platform—An Overview

Rebecca S. Albitz
Bates College Library, ralbitz@bates.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>



Part of the [Library and Information Science Commons](#)

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Rebecca S. Albitz, "Facilitating Content Discovery and the Value of the Publisher Platform—An Overview" (2012). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284315135>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Facilitating Content Discovery and the Value of the Publisher Platform—An Overview

Rebecca S. Albitz, Associate College Librarian for Collection Management, Bates College

Abstract

Libraries invest heavily in content, both through purchase and licensing. This money is wasted, however, if faculty and students are unable to easily locate and use content for research and teaching. Designing and promoting tools to assist in navigating a variety of information sources have been the purview of both librarians and information providers—from the card catalog to sophisticated indexing and abstracting databases to discovery products such as Summon and Primo. Where, however, do publishers fall into the information discovery mix? Should they be investing resources and time into the development of their product platforms, or should they cede content discovery to third parties? Publishers have a vested interest in making their content discoverable through as many paths as possible, and once discovered, publishers want to offer the user additional related content to prolong their stay on the publisher's own site. Librarians also want to provide as many options to our users as possible, rather than investing in a single discovery option.

Introduction

Millions of dollars are spent every year to secure access to content for college and university students, faculty, and researchers. This money is thrown away, however, if the content licensed is not readily located or discovered. All librarians—from collection managers to reference librarians to catalogers—have as a core responsibility connecting users to content. Only the tools available and methodologies employed to do so have changed over the years. From the card catalog to discovery services—each tool offers benefits as well as poses challenges. This evolution in content discovery raises questions about the relative value of one means of discovery over another and whether investment in publisher platform development is a wise use of resources. Specifically, in the age of discovery services, does publisher investment in their own platform make sense? The answer, I would argue, is yes. The following is a brief overview of the evolution of discovery tools, providing the context for the rest of this discussion on publisher platform developments.

The Library Catalog, Indexing and Abstracting Services, and Publisher Websites

Staples of information discovery have been, and I would argue continue to be, the library catalog and the indexing and abstracting (A&I) service. The former leads library users to books purchased, and the latter provides subject-based references to journal literature. One of the values of the local catalog has been that it reflects locally available content. The limitation of catalog-as-inventory, however, is that the universe of accessible rather than owned content is ignored. Now many institutions have broadened the role of the catalog—including records for websites, licensed content, and even electronic books that have not yet been selected for the collection, such as those that are part of a demand-driven access program.

Like the catalog, indexing and abstracting services have evolved dramatically—from those big, green books we used to find *Time Magazine* articles in during junior high to highly sophisticated electronic indexes, many with full text. In the

electronic realm, the platform becomes almost as important as the content. While we are now seeing both platform and content consolidation, at their peak Ebsco, ProQuest, CSA, Silverplatter, OVID, Thomson, FirstSearch, and other platform providers vied for our business, not solely based on content, because many of these information providers offer the same A&I services. To attract more business, platform providers developed sophisticated products with attractive capabilities—controlled vocabulary, natural language searching, subject recommenders, and multiple ways to narrow or expand a search. The more sophisticated, yet easy-to-use the platform, the more librarians are willing to commit to offering multiple resources on that same platform. Thus, functionality becomes a means by which librarians help shape the information platform marketplace.

An additional source of information focused primarily on the librarian market was the publisher catalog, which in many ways is the precursor to the publisher platform. The print catalog showed what a publisher had available for purchase—books, journals, and reference works. One might browse the catalog to get a general idea of what was available, but for many librarians certain catalogs were scrutinized more closely because these publishers were associated with certain subjects. An obvious example would be the American Psychological Association's (APA) catalog, which the psychology selector would have to have reviewed closely. Print catalogs are still available, but other selection tools, such as book vendor websites (GOBI3 for example), electronic review publications, and direct web marketing, now dominate the field of content sales. With these different marketing options available, publisher platforms are not primarily a marketing channel. In some cases these websites might be a place where researchers interested in specific content begin their searching—looking for the latest issue of a journal, for example. This would be particularly true with societies and professional associations—referring back to the APA example. Rarely, however, would a student begin their information exploration on a publisher website. Rather, these sites become destinations, with the journey beginning in an A&I product and

moving through an open URL linking product—the single most important development in the evolution of discovery tools (in my humble opinion).

Cross-Searching and Content Discovery

Even with expanded library catalog content and highly developed A&I platforms, the inability to search across multiple resources at the same time, however, remained a challenge. First to address this were individual platform providers, by developing a way for the user to cross-search all the content an institution licensed from them at once. For example, librarians would have an incentive to license both PsycInfo and Sociological Abstracts on EbscoHost, because both could be searched concurrently. While this solution works, it is highly limiting. It is impossible for a library to license all of its indexing and full-text services on the same platform. Even those libraries that have made a commitment to one primary platform over another will have to utilize multiple platforms, as no single third-party provider is able to offer all content. And, less obviously, if a library configures resources on the same platform to cross-search automatically, usage data for any single product cannot be disaggregated. So, while the sophisticated searching available on a single platform is retained in this cross-searching model, the limitation on content that can be cross-searched makes this solution less than ideal.

The next solution, the federated search, was designed to resolve this problem, allowing the user to search across different products and platforms. This technology, in the form of products such as MetaLib and WebFeat, offered the user the option to select multiple products, enter a search term or terms, and retrieve results without leaving the federated search interface. Unfortunately this option was also less than ideal. The protocols (Z39.50, screen-scraping) used to permit cross-searching were limited both in the search options available and the granularity of results. And, because all platforms could not support these protocols, only a limited number of platforms could be cross-searched. Each time a search was conducted within the federated search system, at least in the case of Metalib, with which I am most familiar, the system had to send the

search criteria out to the individual products' native interface and then retrieve multiple results sets, one for each product searched. Thus, a single search could take minutes to perform. And, more often than not, if the results were too large, the search would hang up, retrieving no results at all. Finally, all of the time and resources platform providers invested in developing sophisticated searching options were lost when a federated search engine was in place. In general, federated searching was slow and clunky and only worked for some products and platforms.

Now we have entered the era of the discovery service—Summon, Primo and Ebsco Discovery Service for example—all of which facilitate searching across platforms and products, retrieving results for books, book chapters, journal articles, newspaper articles, conference proceedings, and unpublished content. Discovery systems provide a Google-like experience, but include only that content librarians have selected and allow for sophisticated refining of search results. And, these searches are quick, because the metadata being searched is aggregated into one knowledgebase. So, is this the ultimate searching solution, eclipsing all others including the publisher platform? Discovery services may seem like a panacea at first glance, and they are very useful to certain user populations. Those who may find discovery services particularly attractive are undergraduates who are just looking for “something” and researchers who are beginning to explore a new subject and do not know the specific language or the information resources associated with that discipline. Searching discovery systems is very simple, and results are easily navigated and refined. But they do present some challenges. Not all information products or platforms are incorporated into every discovery system at the same level, and explaining this shortcoming to users can be difficult. When you offer a single search box, some make the assumption that all products are being searched. But some information providers, particularly those whose primary offering is an A&I product, have chosen not to give discovery services their metadata. This is because, at least in the case of Summon, the discovery service public interface is not behind a firewall. If an institution has

implemented Summon, anyone anywhere can search that instance. They will not be able to retrieve full text, but they would have access to the same metadata for an A&I product as those who have paid for a subscription. Also, discovery services are proprietary products, and some information providers who also offer discovery services are reluctant to work with competitors. So, the metadata for a ProQuest database in Summon (a ProQuest product) will be different than the metadata for the same product in Ebsco's Discovery Service.

The Role of the Publisher Platform

For all of the above reasons, content discovery should not be ceded to discovery services exclusively. So, what is the role of the publisher platform among the myriad of content discovery options available? The first role is to serve as the destination for those looking for full-text content, even when they begin their search in another product. Functionality on the destination platform needs to facilitate linking from outside resources. If this basic functionality does not work, users will become frustrated and avoid using content on the publisher's platform. Basic functionality is also important for those researchers who have a reason to start at a publisher site, which would be because they are looking for a specific title. Again, if this experience is at all frustrating, the platform will be avoided, and licensed content will not be used. The publisher also has a vested interest in making their own platform user-friendly. Once someone has found her way to a publisher's platform, the publisher not only wants to make the target information readily available, but also will want to lead the researcher to other, similar content. So, linking within the platform to articles, book chapters, reference articles, etc., that contain information related to the original content sought is highly desirable. user remains on the platform longer, discovers content previously unknown to them, and is more likely to use and cite this content in their research. Creating this kind of intra-platform linking requires sophisticated metadata and platform functionality, but this investment benefits both the publisher and the user. Finally, librarians do have an interest in publisher platform

functionality. A bad platform, again, will deter use of content. We want nothing more than to have content we have selected be used. And, librarians want to provide as many options to their users as possible, because we cannot predict which path a

researcher will take to find her information. At any given time, a researcher may begin their information search at a publisher site or at Google—we cannot predict. Retaining flexibility is critical.