

DOES DISCOVERY STILL HAPPEN IN THE LIBRARY?

ROLES AND STRATEGIES FOR A SHIFTING REALITY.



Roger C. Schonfeld

Ithaka S+R





Ithaka S+R is a strategic consulting and research service provided by ITHAKA, a not-for-profit organization dedicated to helping the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways. Ithaka S+R focuses on the transformation of scholarship and teaching in an online environment, with the goal of identifying the critical issues facing our community and acting as a catalyst for change. JSTOR, a research and learning platform, and Portico, a digital preservation service, are also part of ITHAKA.

Copyright 2014 ITHAKA. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of the license, please see <http://creativecommons.org/licenses/by-nc/4.0/>.

ITHAKA is interested in disseminating this report as widely as possible. Please contact us with any questions about using the report: research@ithaka.org.

Many of the academic library's traditional roles have come under substantial pressure in recent years. Libraries find themselves grappling implicitly if not explicitly with how to adapt. Discovery is an example of the challenges that libraries face both strategically and managerially as they navigate through large-scale change. It is equally an example of the opportunities they can find by explicitly addressing such changing roles in a well-designed decision-making process that incorporates evidence and judgment.¹

Elsewhere, I have defined discovery as “the process and infrastructure required for a user to find an appropriate item,”² which includes several functions or processes. In exploratory search, one seeks as-yet unknown information on some topic. In known-item search, one seeks to locate a specific information resource already known through previous use, citation, or otherwise. And researchers stay up to date in their field by maintaining current awareness of new findings, theories, methodologies, and so forth. New technologies enable ambient discovery that is geographically aware, search that uses images rather than text, and a variety of other techniques. For all these functions and processes, discovery of scholarly content is very much in flux.

Based on a long-standing and little-changed vision for their role in discovery, academic libraries have in recent years invested in a new generation of discovery services. It is a good moment to take stock of accomplishments and to ensure that wise investments are being made in support of a realistic vision for the library's role.

VISION

The vast majority of the academic library directors who responded to the Ithaka S+R Library Survey in fall 2013 continued to agree strongly with the statement: “It is strategically important that my library be seen by its users as the first place they go to discover scholarly content” (see Figure 1).³ Although that share declined modestly from 2010, library directors seem to perceive continuing value in being seen to serve as the starting point.⁴

- 1 I thank Lettie Conrad, Lorcan Dempsey, Kimberly Lutz, Constance Malpas, Deanna Marcum, and Susan Stearns for reacting to draft versions of this paper. While their suggestions have improved it considerably, the perspective presented here is ultimately my responsibility alone.
- 2 Mary M. Somerville and Lettie Y. Conrad, “Collaborative Improvements in the Discoverability of Scholarly Content Accomplishments, Aspirations, and Opportunities: A SAGE White Paper,” available at <http://www.sagepub.com/repository/binaries/pdf/improvementsindiscoverability.pdf>. I distinguish discovery from discoverability, which encompasses a publisher's or platform's efforts to maximize the usage of its content.
- 3 Matthew Long and Roger C. Schonfeld, *Ithaka S+R US Library Survey 2013* (New York: Ithaka S+R, 2014), available at <http://www.sr.ithaka.org/research-publications/ithaka-sr-us-library-survey-2013>.
- 4 Of course, views about the role of the library can differ tremendously across departments and even from librarian to librarian. With respect to their priorities for discovery, one can anecdotally hear electronic resources librarians looking for improved linking and delivery experiences; collections librarians looking to increase the rate of use of licensed resources; instructional services librarians seeking to improve their information literacy education; and so forth. Distinguishing between the “vision” that each of these individuals may have for improving discovery and the library's organizational vision for its role in discovery is vital.

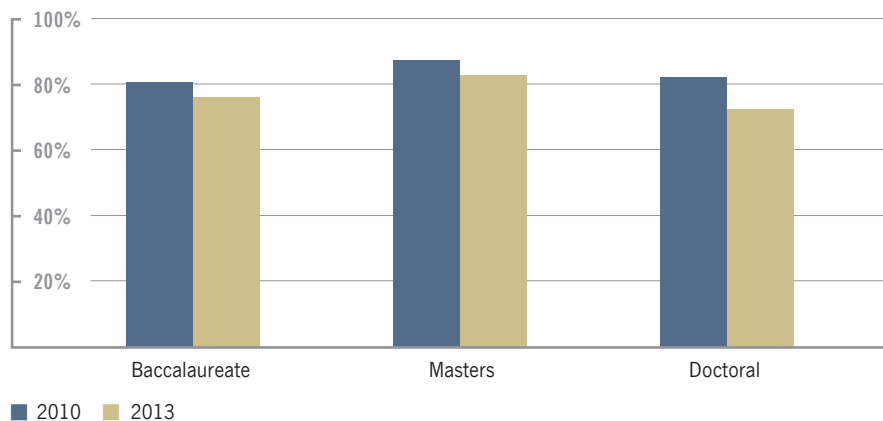


Figure 1: Share of respondents agreeing strongly that “It is strategically important that my library be seen by its users as the first place they go to discover scholarly content.”

There is cause to doubt the sustainability of this vision. Lorcan Dempsey has been outspoken in emphasizing that much of “discovery happens elsewhere” relative to the academic library, and that libraries should assume a more “inside-out” posture in which they attempt to reveal more effectively their distinctive institutional assets.⁵

Even so, serving as the starting point could matter for any number of reasons, political as well as substantive. As libraries seek to influence scholarly communications more directly than ever before, it could be valuable to direct researchers to an open access version of a publication. And as they seek to minimize or at least explicate the information-seeking biases that a discovery service may pose for their users, being positioned to understand and hopefully to influence the starting point resource has real value. Furthermore, being seen to add value to the content being licensed or otherwise provided, and not serving merely as a purchasing agent, is sometimes an important element in the institutional positioning of the library.

Even while they find strategic value in being seen by users as the starting point, directors are less likely to agree that the library is always the best place for researchers at their institution to start their search for scholarly information (as shown in Figure 2, also from the Ithaka S+R US Library Survey 2013). The gap in the share seeing strategic value in serving as the starting point and those that think the library is always best suited in this role suggests that many library leaders may recognize limitations to the vision or failures to date in executing on strategies that could work towards this vision.

5 Lorcan Dempsey, “Thirteen Ways of Looking at Libraries, Discovery, and the Catalog: Scale, Workflow, Attention,” *Educause Review*, December 10, 2012, available at <http://www.educause.edu/ero/article/thirteen-ways-looking-libraries-discovery-and-catalog-scale-workflow-attention>.

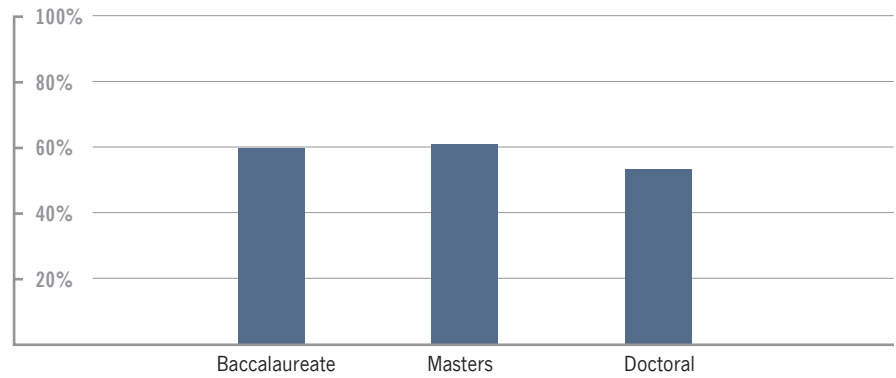


Figure 2: Share of respondents agreeing strongly that “My library is always the best place for researchers at my institution to start their search for scholarly information.”

MARKET SHARE

The academic library has never served 100% of the discovery needs of its users. For current awareness, peer networks have long been recognized for their importance, while in instructional settings faculty members drive student reading through syllabi and reserves. The library’s role has traditionally been focused on exploratory search on a topic and known-item searches, where some 20 years ago the academic library and its librarians had a vital role. Through the public catalog, printed reference resources, the reference librarian, and browseable stacks, academic libraries had built and gathered a valuable set of infrastructure, staffing, and services in support of search. The library was able to serve a high share of the search needs of its community.

In recent years, new entrants have built many different types of search and discovery services that have reduced the centrality of the library’s services. To be sure, library services have innovated: reference resources have systematically moved online, as A&I services and secondary publishing platforms; journal and ebook platforms offer powerful full-text search; and reference assistance is provided online and through improved instructional offerings. But search has moved to the network level, and whether it is through Google’s Search, Scholar, or Books services, Wikipedia, or a variety of other tools, a higher share of academic discovery than ever before is routed around, rather than through, the library.

Does the library’s reduced market share for academic search matter? Many might argue that the library is not competing for market share anyway, or that search has become commoditized and libraries should retreat from this role systematically.⁶ Still, the decline in library market share for search conflicts with the vision that directors expressed in their responses to our survey, as well as the investments that their libraries have been making in support of this vision. If market share has declined, but resource and staff expenditures have not declined as much, then the library could be seen as spending relatively more; possibly reasonable for any number of reasons but also possibly suggesting that resource allocations are coming into misalignment.

6 Dale Askey for example has argued that “Google won the discovery wars years ago, and nothing we can do is going to change that.” “Giving up on discovery,” Taiga Forum, September 17, 2013, available at <http://taiga-forum.org/giving-up-on-discovery/>.

...a higher share of academic discovery than ever before is routed around, rather than through, the library.

SYSTEMS

To date, the most important strategy for libraries seeking to realize this vision of serving as the research starting point (and recapture declining market share for exploratory search) has been the introduction of various systems designed to bring together as high a share as possible of the library's collections into a single search interface. Federated search systems were an antecedent to the index-based discovery services that have achieved a market penetration of roughly 70% at baccalaureate and master's institutions and roughly 85% at doctoral universities. The vast majority of the libraries that have licensed such services have made them the default search on the library homepage, thus emphasizing the role of these services as the starting point for research needs.⁷

There are arguments back and forth about the extent and nature of the impact that indexed discovery services are having on search behaviors. Providers of these services have presented findings suggesting that individual libraries are handling much more search traffic than was the case previously, and that many content providers, especially smaller ones, are seeing significant increases in content accesses.⁸ Some have wondered if the arresting of the trends away from the local library interface and towards electronic research resources, as found in the Ithaka S+R US Faculty Survey 2012, may represent the beginnings of real impact resulting from the indexed discovery services.⁹ At the same time, some of the medium to large content platforms, such as IEEE, JSTOR, and SAGE, report that a vastly greater share of content accesses derive from Google-sourced sessions than from those sourced from indexed discovery services.¹⁰ Finally, recent research suggests that following the implementation of a discovery service, usage of some publishers' journals increases more than those of other publishers, with important differences from library to library and among the various discovery services.¹¹ Unfortunately, little is known directly about the usage of these discovery services themselves, although there is hope that a common usage data framework, still only in its infancy, will improve matters.¹²

7 Long and Schonfeld.

8 See for example Jenny Walker, "Advancing Discovery Services for Libraries," presentation at the Fiesole Retreat, April 2012, available at http://www.casalini.it/retreat/2012_docs/walker.pdf.

9 Eddie Neuwirth, "The (Possible) Impact of Discovery -- Results from the Latest Ithaka S+R Faculty Survey," available at <http://www.proquest.com/blog/2013/the-possible-impact-of-discovery-results-from-the-latest-ithaka-sr-faculty.html>.

10 Personal communications. Due to authentication protocols, there are anomalies in how traffic sourcing from discovery services is tracked, which may lead to their underestimation in some cases, although it is not believed that this affects the relative pattern. It is not known if similar patterns obtain for other scholarly resources.

11 Michael Levine-Clark, John McDonald, Jason Price, "Discovery or Displacement? A Large-Scale Longitudinal Study of the Effect of Discovery Systems on Online Journal Usage Publication," *Insights: the UKSG journal*, 27:3 (November 2014), forthcoming.

12 Usage statistics are covered in Section 3.3.4 of this recent report: *Open Discovery Initiative: Promoting Transparency in Discovery*, NISO RP-19-2014, issued June 26, 2014, available at <http://www.niso.org/workrooms/odi/>.

Attitudinal data from library directors, which may not reflect actual dynamics but certainly their impressions, is shown in Figure 3. When they were surveyed, respondents indicated that indexed discovery services had broadly improved exploratory searching, but that this was less the case for known-item searching.

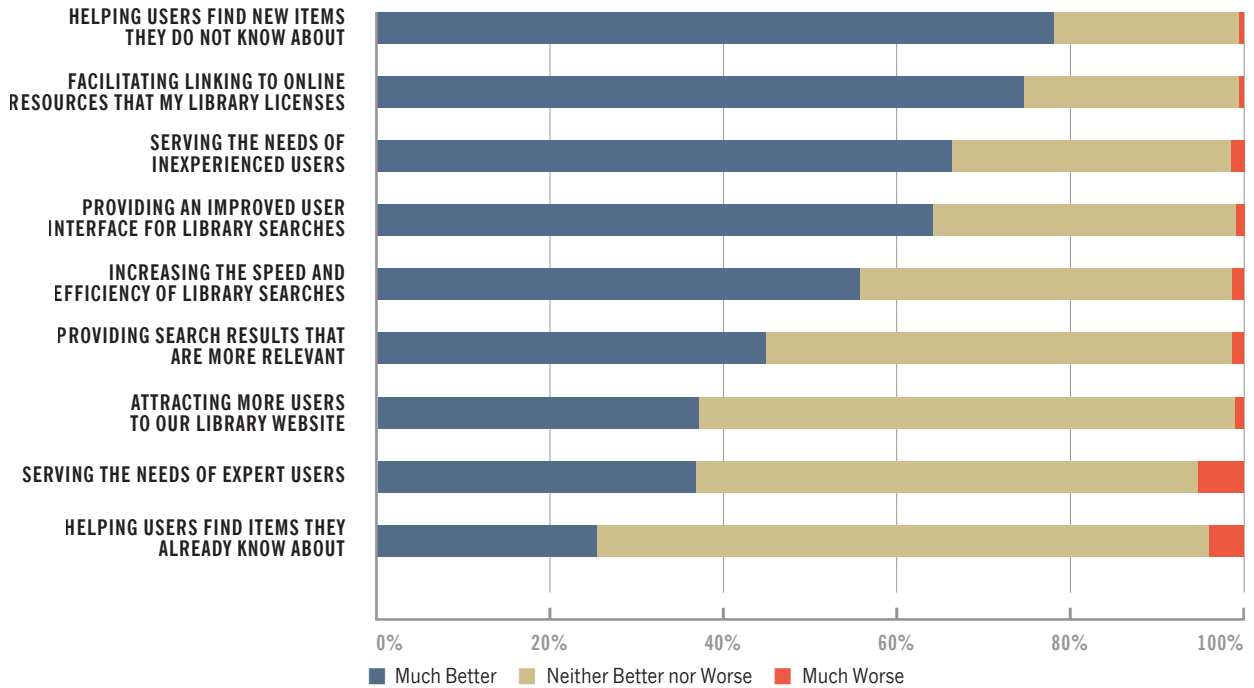


Figure 3: “To what extent do you think that your index-based discovery service has made your users’ discovery experience better or worse in each of the following areas?”

...this is a good moment for academic libraries to step back to reconfirm (or reconsider) their vision for discovery...

Notably, only a minority of responding library directors believes that the discovery services have helped bring more users to the library website, a finding that might be juxtaposed with their belief in the strategic importance of being seen as the starting point. While there are surely other ways to measure the user perception of the library as starting point, many libraries’ vision for the role of indexed discovery services include recapturing market share from Google and other third parties.

Usage patterns and perceptions of them surely lag behind the availability of new features in the discovery services, and there are many indications that innovation in these services’ offerings will continue. Still, this is a good moment for academic libraries to step back to reconfirm (or reconsider) their vision for discovery, to ensure that their visions connect with information-seeking practices and preferences, and to determine whether they have a viable strategy in place, beyond the choice of systems, to achieve their vision.

PRACTICES AND PREFERENCES

The single search box is highly appealing as a library strategy for discovery, not least because of the great success that Google has achieved through this interface. But in developing their strategies, libraries benefit from unpacking in turn community practices and preferences, and what exactly is meant by a “single” search box.

Community practices and needs vary tremendously. Among faculty members, discrete practices emerge for known item searching, exploratory searching, and current awareness, with discipline serving as an extremely important variable for all three of these discovery cases. Figure 4 shows that for known item searching, the importance of the library website compared against scholarly databases or search engines may be nearly reversed from sciences to humanities.¹³

Community practices and needs vary tremendously.

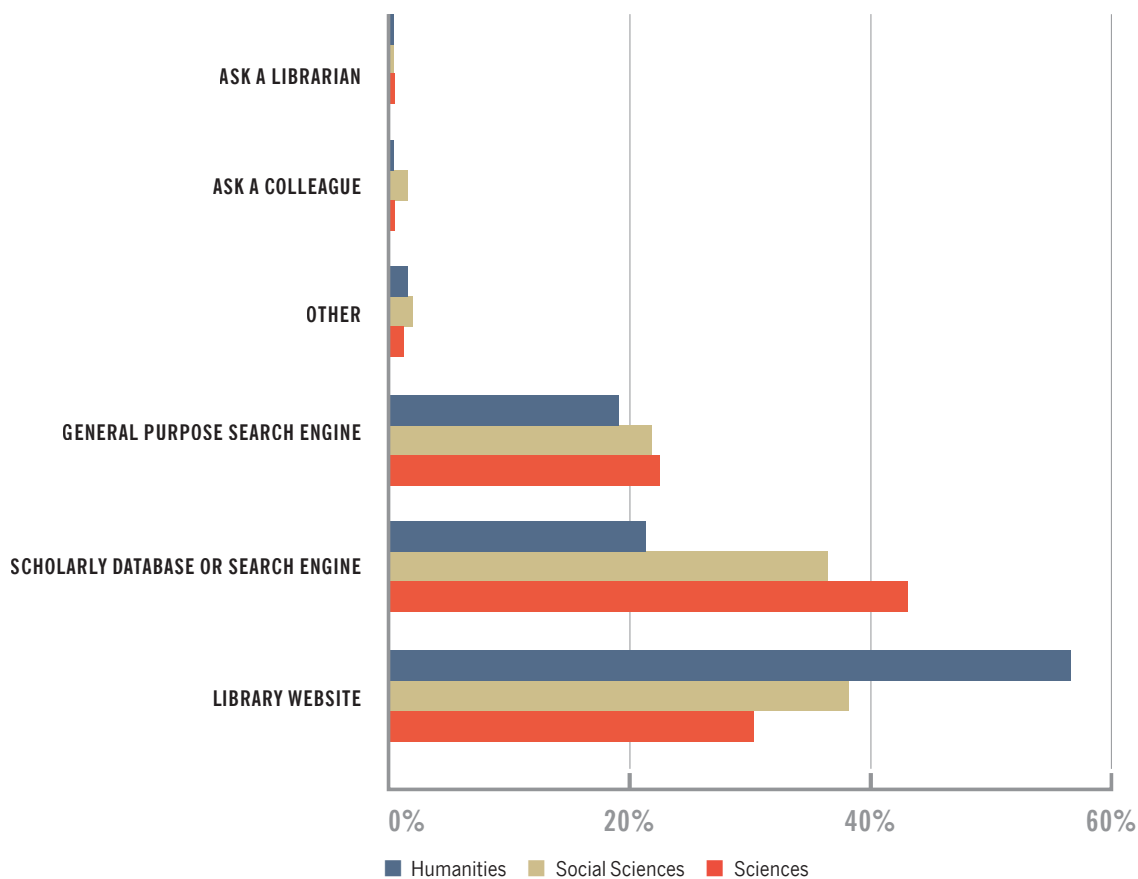


Figure 4: “When you try to locate a specific piece of secondary scholarly literature that you already know about but do not have in hand, how do you most often begin your process?”

And while the practices of faculty members may tend to vary by discipline, among students there is significant variance by level of experience. First-semester freshmen are remarkably different from seasoned master’s students.

13 Ross Housewright, Roger C. Schonfeld, and Kate Wulfson, *Ithaka S+R US Faculty Survey 2012* (New York: Ithaka S+R, 2013), available at <http://www.sr.ithaka.org/research-publications/us-faculty-survey-2012>.

With respect to both faculty members and students, participants in Ithaca S+R’s local surveys are beginning to indicate tremendous institutional diversity. Parallel local surveys ran in 2013-14 at Indiana University and University of North Carolina at Chapel Hill—both large research-intensive public institutions—illustrate that university context may in some cases matter more so than do other factors. Both for undergraduate students (Figure 5) and faculty members (Figure 6), a significantly higher share at UNC report starting at the library website or online catalog than do so at Indiana University.¹⁴

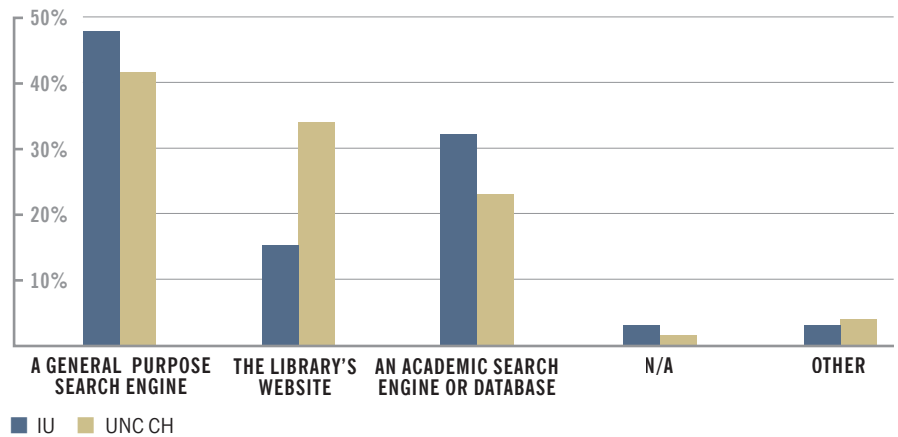


Figure 5: “Please think about your most recently completed research paper or project. Which of the following starting points did you use to begin your research?” (undergraduates)

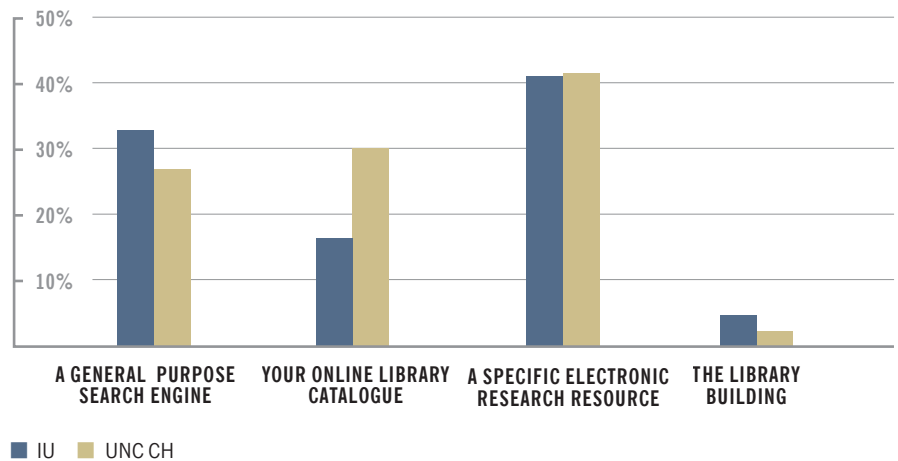


Figure 6: “Below are four possible starting point for research in academic literature. Typically, when you are conducting academic research, which of these four starting points do you use to begin locating information for your research?” (faculty members)

While greater analysis of actual usage patterns would bolster this argument, a strong hypothesis can be advanced that practices and perhaps also needs vary between faculty members and students, among faculty members by discipline, and among institutions as well.

14 I thank Andrew Asher of Indiana University Bloomington and Heather Gendron of the University of North Carolina at Chapel Hill for making it possible to include their universities’ findings in this paper.

The concept of a single search box also requires critique.

...many researchers would benefit tremendously from increased personalization in discovery.

SEARCH BOX

The concept of a single search box also requires critique. While Google offers a merged search, it builds this from stand-alone verticals such as Images, Websearch, Books, and Scholar, each of which remain available on a stand-alone basis. Perhaps more significantly, many single box consumer services have increased the level of personalization that they offer, through preference-setting and through data-gathering such as search history, location, and more. It is also possible to search not only online materials but also personal collections in services such as Evernote through easily integrated browser extensions. Google's single search box thus may seem powerfully simple, but it actually offers a highly differentiated experience for each user.

Given the amount of data being gathered by consumer search services and social networks, we might assume that the more signals available to an advertising or discovery algorithm the higher quality the search results. By contrast, libraries have not pressed their developers and vendors to personalize their services. While rightfully cherishing the value of privacy even as so many of us gleefully forego it in consumer services like Facebook, librarians are also beginning to ask how values change in tension with societal shifts.¹⁵ Some indexed discovery services allow researchers to set preferences such as their field of study, and there is every reason to believe that many researchers would benefit tremendously from increased personalization in discovery.

Whether the library search box, perhaps powered by indexed discovery services, grows more powerful, many observers agree that no homepage can hope to be the single starting point for the content it contains. The recent New York Times strategic review emphasized that even such an internationally known and trusted brand could not sustain the homepage as the wrapper for its journalism.¹⁶ Will this prove to be the case for academic search? If so, what are the implications for a vision for the library's role in discovery?

ALTERNATIVES

Discovery has occupied a growing amount of systems resources and attention in recent years in academic libraries, with focus principally on exploratory and known-item searching. Academic libraries are shifting strategy, reorganizing staff, and licensing or building new library systems, to a great degree in support of a vision that the library has a central role to play here. Is this vision the right one for the academic library?

In a different discovery use case, scholarly current awareness in one's field, the search box has never been the right interface even if a large index could prove to be quite valuable. Researchers' tactics for maintaining current awareness draw especially strongly on peer networks and academic conferences; for humanists,

15 See for example Barbara Fister, "What's So Sacred About Privacy?| Peer to Peer Review," *Library Journal*, May 29, 2014, available at <http://lj.libraryjournal.com/2014/05/opinion/peer-to-peer-review/whats-so-sacred-about-privacy-peer-to-peer-review/>.

16 The New York Times's internal report on "Innovation" was leaked. It is a fascinating read not only about digital and hybrid strategy but also about organizational structure. It is available at <http://www.scribd.com/doc/224608514/The-Full-New-York-Times-Innovation-Report>.

Can libraries step forward to play a greater role in current awareness?

book reviews and publisher announcements are notably important as well. Little effort has been made to improve the fragmented eTOC alerts that seem to be the best publishers can offer, for example hosting the peer discovery experience in an online social networked space or bringing book reviews out of their confines in journals.¹⁷ But new services that offer and in some cases combine social discovery, article management, and professional profile, are addressing these gaps, as are services like Google Scholar's My Updates.¹⁸ Can libraries step forward to play a greater role in current awareness? Should they do so?

Personal collections are frequently ignored but are growing in importance to scholars and pose challenges in terms of their discovery. Some of the best discovery services offer options, either directly or through scripts, to search your Kindle library, your Dropbox and Evernote files, the materials on your hard drive or phone or tablet, and so forth, right alongside a global search. Ultimately, even when conducting a search of the scholarly literature, won't users wish to be connected with access to the Kindle copy they've purchased rather than the library-licensed e-book provider? Seamless discovery of personal collections alongside institutional collections might be extremely attractive.

Or, is there a more integrated vision for discovery that libraries and their systems providers and primary and secondary publishers can collectively pursue? While academic communities are understood as institutionally affiliated, what would it entail to think about the discovery needs of users throughout their lifecycle? And what would it mean to think about all the different search boxes and user login screens across publishes and platforms as somehow connected, rather than as now almost entirely fragmented? It could be that the biggest advantage of the suite of services that Google offers is that it cuts across use cases, the user lifecycle, and content platforms, more so than its single search box interface. Libraries might find that a less institutionally-driven approach to their discovery role would counterintuitively make their contributions more relevant.

Libraries might find that a less institutionally-driven approach to their discovery role would counterintuitively make their contributions more relevant.

Alternatively, it might just be that free searching such as that provided by Google or Scholar is effective enough that the library can walk away from making investments of its own. The prospect for the library to cede discovery as a function and rely on Google is becoming the baseline service level or "control group" against which to test not only the value but also the cost of any other discovery strategy the library might choose to pursue.¹⁹

17 On the latter point, see for example Oona Schmid, "Faster and Cheaper: Can a Digital-Centric Workflow Transform the Book Review?" *Ithaca S+R Issue Briefs*, August 27, 2014, available at <http://www.sr.ithaca.org/blog-individual/faster-and-cheaper-book-review>.

18 For recent coverage of some new developments in current awareness, including social discovery, see Elizabeth Gibney, "How to tame the flood of literature," *Nature* 513, 129–130 (04 September 2014), available at <http://dx.doi.org/10.1038/513129a>.

19 Even if this formulation is seen as refreshingly provocative, we mustn't blind ourselves to the unknown sustainability of Google Scholar in particular.

DEFINING SUCCESS

Libraries have promoted their web sites as the starting point for discovery and have invested in systems to bring their myriad collections into a single search interface. Has this investment been worth it? This is not a hypothetical question, but one that academic libraries can answer today through rigorous review. If success were defined, ideally quantitatively, it would be possible to assess whether this strategy has succeeded. For example, have the share of content accesses derived from library discovery services grown by a certain amount? Is the number of unique visitors per week to the library homepage growing by a certain level? Is a sufficiently high share of the content accesses to key e-resources sourced from the library's discovery services? Has the share of the library's community that values the library's role in support of discovery grown? With metrics such as these—ideally defined when the strategy was implemented but even retroactively as needed—libraries can assess the success of a strategy relative to the staff and resources expended in pursuing it.

As a good steward of its institution's resources and its community's needs, the library should in this way determine whether its strategy is successful, and if not how to adapt. A library that determines the strategy to be succeeding can take pride and watchfully monitor with an eye towards continuous improvement. But a library that determines the strategy is not succeeding should not necessarily double down on the vision with an alternative strategy. While it is possible that an alternative strategy, perhaps one that embraces deep personalization or current awareness, could succeed, it is also possible that the vision—that the library has an essential role to play in facilitating discovery—could itself be flawed.

Diagnosing success or failure is a useful exercise, whatever big issue is being addressed. Libraries cannot be afraid of data that indicates a course correction is necessary. The ultimate goal is to expend resources in alignment with a vision that can work and a strategy that can succeed. An evidence-based approach is essential.