Semantic Studios

Search & Discovery

A Roadmap of Possibilities

for the Future of the Harvard Library

Harvard Library

Submitted to Mary Lee Kennedy

November 8, 2012
# Table of Contents

**Analysis** .................................................................................................................. 3  
  Pervasive Problems.................................................................................................. 4  
  Emerging Solutions .................................................................................................. 6  
  Flexible Facets .......................................................................................................... 9  

**Recommendations** ................................................................................................... 10  
  Search as a Service .................................................................................................. 10  
  Benefits of Action .................................................................................................... 12  

**Next Steps** .............................................................................................................. 14
Analysis

In recent years, our ecosystems of information, research, teaching, and learning have been subject to dramatic, ongoing technology-driven change. In response, Harvard has begun the transformation towards a more integrated library organization that can address the challenges of today and tomorrow. The Harvard Library has established a governance framework, begun operations, and launched a new portal that serves as a gateway to library collections, facilities, and services. These are exciting first steps, but if the Harvard Library is to realize its potential, much remains to be done.

Search and discovery, one of the library’s most critical strategic challenges, is exposed as a problem on the portal. Users are inclined to ask “why can’t we search all of the library’s collections at once?” and “why do I need to know in which database my result lies before I begin to search?” These questions defy simple answers.

[Figure 1. To search, users must first choose from dozens of databases.]
In fact, Harvard has been aware of the search and discovery challenge for several years, as evidenced by the 2009 Task Force report.\(^1\)

Through extensive consultation with the community, the Task Force has heard from students, librarians, faculty, and other interested parties how singularly difficult it is to find the resources which, when discovered, prove invaluable.

And, the 2012 Harvard Library Portal Online Survey showed that students, faculty, and staff perceive search and discovery as the single most important problem.

“I think it’s better to have everything in one searching system.”

“(We need) better discovery tools that integrate electronic articles with books and manuscripts.”

“(Harvard should) prioritize search over all else.”

In part, progress has been stifled by the technical and legal obstacles inherent in searching across local and licensed content and metadata from myriad vendors and sources. To make things worse, Harvard lacks a clear vision for search and discovery, for while there’s obvious demand for a “single search box” that works like Google, expert searchers prefer the powerful, native database interfaces that they’ve learned to love, and they worry about the “dumbing down” effect of a lowest-common denominator search.

**Pervasive Problems**

Of course, Harvard is not alone in wrestling with this conundrum. For more than a decade, academic and public libraries worldwide have failed to solve this problem. As a consequence, faculty tend to stick with familiar databases (thereby missing opportunities to conduct cross-disciplinary searches and learn about new sources), while students often opt for the fast, easy solutions of Google and Google Scholar, as this study shows.\(^2\)

Regardless of all the time and effort libraries put into providing a variety of research tools and resources on their websites, the literature suggests that students still prefer to start their research using Google or some other form of search engine…It is clear that there is an overwhelming preference for easy to use, familiar search tools that transcend education level, discipline of study, and student demographics.

---


2 Discovery Layers and the Distance Student, Jessica Mussell (2012).
While Google Scholar has obvious strengths, it’s certainly not without weaknesses:

- No “advanced search” functionality.
- Limited, inaccurate metadata.
- Inconsistent coverage across disciplines.
- Not customizable or interoperable.
- No transparency (e.g., coverage, algorithms, usage, monetization).

This last point should be of serious concern to any institution that plans to rely upon Google Scholar as part of its search and discovery strategy. We’re not sure what is and isn’t being searched. We don’t know how or why it ranks some results higher than others. We learn nothing about the needs and behaviors of our searchers. And, we’re in the dark regarding the financial realities and sustainability of Google Scholar.

It gets worse. As shown in a recent study\(^3\) by Project Information Literacy, students are entering the workforce without the research skills that are expected and required:

Employers claimed that college hires rarely conducted the thorough research required of them in the workplace.

At worst, some college hires solved problems with a lightning quick Google search, a scan of the first couple of pages of results, and a linear answer finding approach.

Employer: “I had new graduate hire who only searched for papers on Google. I said, you’re missing things, you need to use PubMed, and he responded, ‘Well, I did this quick search, and that’s what I got.’ But that’s not good enough.”

By failing to provide a desirable search solution, libraries are letting down their students and faculty. They also risk damaging themselves irreparably.

A 2009 Ithaka S+R Faculty Survey\(^4\) showed a decline in the perceived importance of the gateway function (search and discovery) and a corresponding increase in the perceived importance of the buyer role (paying for books, journals, databases).

\(^3\) Project Information Literacy Research Report: Learning Curve by Alison J. Head (2012).

The report cautions that “the academic library is increasingly being disintermediated from the discovery process, risking irrelevance in one of its core functional areas.”

**Emerging Solutions**

To address this problem, academic libraries and library vendors have been working towards a “single search box” solution, with varying degrees of success. In order to understand the mix of possible solutions, it’s important to separate what you see on the front end (*the interface*) from what’s happening on the back-end (*the index*).

On the back-end, a key development is the transition from a federated model towards an aggregated search index:

- **Federated Model.** The user’s query is sent to several search systems, and results are then consolidated into a single display. The primary problem with this approach is that it’s prohibitively slow. Users lack the patience to wait for results.

- **Aggregated Search Index.** Metadata from multiple sources and types of content (*e.g., books, journal articles, manuscripts, images*) is integrated into a single database or index. The user can query a single search box and immediately receive an integrated result set. This enables iterative, interactive search behavior.

Libraries and vendors have recognized that the federated model is too slow, so most development efforts are now targeted towards the aggregated search index.
On the front end, there are a couple of user interface models being used by libraries for search and discovery, both of which feature a single search box. One option, known as the “bento box” style, segments search results by content type and source.

[Figure 3. The “bento box” model for search results.]

The “bento box” approach makes the user aware of the major content types and sources, offers a glimpse of results from within each index, and invites them to go directly to the appropriate catalog or database to continue their search.

A problem with this approach is that it’s dramatically different to the search user experience with which people have become familiar. The single search box promises the simplicity of an Amazon or Google search, but then users are faced with a complex interface that’s difficult to parse. Also, this model typically requires a multi-step search process (i.e., enter query, parse results, select source, parse results) that feels unwieldy. As a result, experienced users are likely to go directly to familiar sources when possible, and rely upon the “bento box” search only as a choice of last resort.
That said, this model has been adopted by several universities including NCSU, Stanford, Dartmouth, University of Virginia, and Columbia.

A second option, the faceted navigation model, returns an integrated result set within the main body of the page, and an interactive list of facets and values on the left. A major advantage of this model is that users can search as they normally do, by entering a query and scanning a single list of results. But, if they don’t find what they need right away, they can look to the left, where the faceted display does two things:

1. The facets/values serve as a custom map of search results, helping people to understand the size and scope of matching results.

2. Each facet value serves as a simple next step, helping people to narrow or refine their queries. And, by taking several simple steps, users can construct what in the past would have required a sophisticated Boolean query.

[Figure 4. The faceted navigation model.]
It’s because of these advantages that the faceted navigation model has become nearly ubiquitous in e-commerce and extremely common on most types of large websites. In academic libraries, the faceted navigation model has recently been adopted by Cornell, Duke, McGill, Northwestern, and the University of Washington.

Initial results of these implementations have been positive. For instance, in a usability study focused on Summon, a “web-scale discovery” tool that combines an aggregated index with faceted navigation, the single search box was used for 80% of the assigned common library search tasks. Users favored the new search tool over the traditional library catalogs and databases. Even better, as evidenced by a different study, the use of full-text online content dramatically increased in the year following implementation of Summon. And, the librarians found they could focus their instruction less on choosing and using a database or catalog and more on refining a search, research as an iterative process, and other high level search skills.

Flexible Facets

The idea of a single search box with an aggregated index and faceted navigation is sometimes met with resistance, particularly within academic libraries. Some concerns are valid, while others result from a lack of understanding. For instance, some faculty and librarians worry about losing the ability to construct sophisticated queries with their familiar native database interfaces. This concern is valid but can be addressed.

First, adoption of a unified search solution does not require the abandonment of native database interfaces. Use of the discovery layer is optional, so expert searchers may continue to go directly to their favorite catalogs and databases. And, for novice searchers, the discovery layer can be designed not only to afford direct access to content, but also to point people to the native catalog and database interfaces.

Second, the faceted navigation model has more power and flexibility than many faculty and librarians may realize. While it’s true that a top-level search across all catalogs and databases cannot fully leverage the available metadata (since it may not be common or interoperable across collections), once a user selects a collection or content type, unique metadata can be brought into play. These “adaptive facets” are used widely on Amazon, where unique facets such as Megapixels appear upon selection of the Digital Camera category. Similarly, it’s possible to design an advanced search interface that adapts its filters based upon a user’s selection of sources.

5 How Users Search the Library from a Single Search Box by Lown, Sierra, Boyer (2013).

6 The Impact of Serial Solutions’ Summon on Information Literacy Instruction by Stephanie Buck and Margaret Mellinger (2011).
Recommendations

Harvard Library should invest in the development of an information architecture for a universal search and discovery service that improves access to local and remote sources from multiple points of origin. This process should integrate user research (i.e., consultation with students, faculty, librarians), stakeholder interviews, and collaboration with working groups (e.g., content, metadata, search, governance) and related initiatives in such areas as digital asset management and interoperability. The goal is to develop a high-level, user-centered information architecture that serves as a map for design and a boundary object for ensuing discussions about implementation.

Search as a Service

It’s vital to recognize that the universal search and discovery service will not simply be concerned with the user experience that starts at the library portal, nor will it focus solely on providing access to traditional library catalogs and licensed databases.

[Figure 5. Search as a service allows for multiple starting points and sources.]
There are multiple points of origin, places where questions get asked and searches begin. For instance, the universal search service should make it possible for faculty and librarians to embed custom search widgets at a course-specific level within LibGuides and learning management systems. Evidence suggests that when you integrate trusted information resources into the student workflow, they’ll tend to rely on those resources rather than going to Google or Google Scholar by default.  

Also, the number and diversity of sources continues to grow, particularly given the rise of open access disciplinary and institutional repositories. A universal search solution must have the flexibility to integrate new local and remote sources of content over time. Similarly, the scope of search should not be limited to content, but should also include library facilities and services. Studies show that when offered a single search box, users will search not only for content but also for specific catalogs and databases, as well as library locations, hours, events, and services.  

Regular review and analysis of the search logs together with implementation of a “Best Bets” capability will ensure answers for all common queries. It’s no longer good enough to blame the user for searching in “the wrong box.” Universal search must provide access to all important, relevant sources of information.

---

7 Embedding LibGuides into Course Management Systems by Stephanie Brown (2012).

8 How Users Search the Library from a Single Search Box by Lown, Sierra, Boyer (2013).
Since it’s impossible to imagine or act upon all of the possible applications, the search service must be open and interoperable. During the past decade, we have seen the tremendous innovation that’s been unleashed by the open APIs and developer platforms of Amazon, Apple, Google, and Facebook. Harvard should do the same.

In a similar vein, an open source search and discovery platform will not only enable Harvard to customize its solution(s) over time, but might also enlist a wider community of academic libraries in the ongoing development and improvement of the platform. This type of commitment to openness, transparency, and interoperability will unleash the creativity of the Harvard community and beyond.

**Benefits of Action**

Harvard can expect a substantial return on its investment in universal search and discovery. A tool that delivers both a simple, fast search across collections and a powerful, deep search within collections will serve as an appealing alternative to Google and Google Scholar. By drawing students into a more iterative, interactive search process that employs “maps of knowledge” to afford deeper understanding and the use of more diverse sources and content types, universal search will contribute towards improved scholarship, learning, and literacy. Similarly, cross-disciplinary search features will help faculty to discover and explore new sources.

On the publishing side, a unified search platform will enhance the visibility of faculty publications (*e.g.*, through integration of DASH) as well as Harvard websites and apps. It will also make it easier for faculty and librarians to integrate information resources and discovery capabilities (*e.g.*, embeddable, customizable search widgets) directly into disciplinary LibGuides and course-specific learning management systems.

Harvard as a whole will benefit from increased efficiencies in design, technology, and licensing. Also, the unified search and discovery service will improve the institution’s ability to utilize existing sources and services via enhanced interoperability.

Finally, Harvard will demonstrate its commitment to the future of the academic library as a source of intellectual value and vitality that advances institutional goals around research, teaching, learning, and literacy. By centralizing rather than outsourcing this function, Harvard will regain access to the behavioral data necessary for ongoing evaluation and improvement. And, Harvard will retain maximum flexibility to adapt its platform in response to emerging challenges and opportunities.
As evidenced by this disturbing list of trends, there’s no question that today’s academic libraries are faced with many challenges:\(^9\)

1. Collection Size Rapidly Losing Importance
2. Traditional Library Metrics Fail to Capture Value to Academic Mission
3. Rising Journal Costs Inspiring Calls for Alternative Publishing Models
4. Viable Alternatives to the Library Now Boast Fastest Growth and Easiest Access
5. Demand Declining for Traditional Library Services
6. New Patron Demands Stretch Budget and Organizational Culture

There is certainly a risk that academic libraries will lose external support and internal staff and skills and infrastructure to an extent that’s undesirable and unrecoverable. However, this time of change also affords new possibilities for the library. While a full account is outside the scope of this report, the key areas are worth mentioning:

• Unified Search & Discovery
• Information Literacy (*research skills for the 21\textsuperscript{st} century*)
• Cross-Channel Design (*bridging physical and digital experiences*)
• Library as Place (*knowledge creation and collaboration*)
• Library as Platform (*open access, open source, interoperability, edX*)

While each of these areas holds great promise, it’s hard to imagine a vibrant future for the academic library without the core intellectual foundation of search and discovery. Rather than outsourcing the gateway function and becoming further disintermediated from the discovery process, the Harvard Library should provide leadership by defining and implementing a vision for a universal search and discovery service. Achieving success will not be easy, but it will be worth it.

---

\(^9\) Redefining the Academic Library by The Education Advisory Board (2011).
Next Steps

Harvard Library should commit to development of an information architecture blueprint for a universal search and discovery service that improves access to local and remote sources from multiple points of origin. The process should include:

- User Research (*students, faculty, librarians*).
- Stakeholder Interviews (*faculty, librarians, administration*).
- Working Group Meetings (*content, metadata, search, governance*).

Consultation with teams working on related initiatives such as interoperability and digital asset management will also be necessary to ensure strategic alignment.

This process will result in an Information Architecture Blueprint that integrates narrative, diagrams, and wireframes (*sketches of major interface types*) into a compelling portrait of universal search and discovery for the Harvard Library. This architecture will serve as a map for design and as a boundary object for ensuing discussions about phased implementation (*e.g., roles, technologies, processes, standards, roadmaps*).

The need for a holistic information architecture for Harvard Library has been recognized for some time, as shown by these 2009 Task Force recommendations:10

The Task Force believes that Harvard must develop a robust, shared information architecture to guide future development and to orient investments in innovative projects…The Task Force recommends that the libraries formulate a vision for mapping the online resources and coordinating technologies available across Harvard, including integrating the libraries’ resources with other online entities such as course platforms, iCommons, and research collaborations.

Just as the carefully planned and often glorious architecture of the library buildings can carry tremendous symbolic weight and increase the functionality of its mission, so too does this reflect upon its online architecture.

Now that the Harvard Library has established a governance framework, begun operations, and launched a new portal, the time is right to develop this information architecture, so that future projects can be guided by a sensible, holistic vision for search and discovery.

---