



# American Library Association Search Assessment

Duane Degler

Search Strategist & Principal, Design for Context

Daniel Funk

Senior Search Engineer

## Contents

[Executive Summary](#)

[Introduction](#)

[Now: The Challenges](#)

[Organizational Content Challenges](#)

[Disparate organizations and identities](#)

[Disparate audiences, experience, and needs](#)

[The “ALA voice” is diffuse](#)

[Content managers have many responsibilities and your members have high expectations.](#)

[Problems with Search](#)

[Problems with the Previous Drupal-based Search](#)

[Problems with Google Custom Search](#)

[Looking Forward: Solution Options](#)

[Overall Attributes of an Effective Solution](#)

[Benefits of a Complete Approach to Search](#)

[How Search Supports ALA Goals and Expectations](#)

[Capabilities Over Short-term, Mid-term, and Long-term](#)

[Technical Options](#)

[Technical option 1: Google Search \(Free\)](#)

[Technical option 2: Paid Third Party Service](#)

[Technical option 3: Standalone Search](#)

[Additional Capabilities and Technologies](#)

[Design and Content Requirements](#)

[Recommendations](#)

[Option 3 Offers Long-Term Benefits](#)

[The Potential Role of Partner Organizations / Companies](#)

[Consultant Biographies](#)

[Daniel Funk, Senior Search Engineer, OpenSource Connections](#)

[Duane Degler, Search Strategist & Principal, Design for Context](#)

[Appendix A: Details of Capabilities Requirements](#)

# Executive Summary

## Search Challenges

**ALA.org search isn't working effectively for its users, its content providers, or ALA as a whole.** Even with recent changes, it is not possible to *"find what I need on page 1"* of most search results. In addition, there are a large number of high value ALA sites that are not available to search, because everything that is not under some part of the ALA.org URL is not included. As a result, search doesn't bring value to the surface or drive traffic to revenue-generating and user-supporting content. For content managers, there is a tension between the desire for completeness in capturing metadata (particularly topical information about the content) and the time and ability of the staff to fully populate the detailed metadata.

**Users have a very wide array of needs and interests, and the association has a wide array of sites and URLs that aim to meet those user needs.** But this complex matrix of people and content needs a sophisticated approach to search. The recent change to Google Custom Search does not address user needs -- search quality is not optimal. This is due to the lack of appropriate metadata, no ability to control for content importance, and no ability to provide additional filters to manage large volumes of content. Prior to the change, the Drupal-based solution was not adequate to tune for results ranking and performance. Both solutions fail to include a significant portion of ALA's important content from related sites or documents.

The upcoming -- and welcome -- move to a new ALA.org web design and mobile focus will increase the demands on search. Given the large volume and diversity of ALA content, the commitment to accessibility and recent improvements will also increase demands on search.

## Search Solutions: 3 Options

This report outlines three technical options as possible cost-effective solutions: Google Custom Search, an alternative paid third party service, or a standalone open source implementation. Each of these are described with a simple outline of what they could offer to ALA users, and are associated with a magnitude of effort, as well as a range of hours and cost to implement.

Along with these three solutions, the report outlines additional technical capabilities and design activities that increase quality and support long-term maintainability.

The table on the next page provides comparison of the magnitude of effort for each of the options described in the report. Please note that the actual timeframes for implementing each approach will vary based on the number of people assigned to work the implementation effort and the balance of internal/consultant responsibilities. Please also note that the ranges provided are intended to provide ballpark estimates of implementation efforts and would vary, depending on the specific functionality targeted as part of any implementation effort. A more detailed proposal targeted to provide a full featured site search for ALA is currently underway and will be provided in February.

Option from Assessment Report	Option Description	Magnitude of Effort	Range of Hours	Range of Cost (in USD)
Option 1	Google Search	1	60 to 180 hrs.	11,700 to 35,100
Option 2	Paid 3rd Party Service-Hosted Search (Swifttype)	3	120 to 240 hrs.	23,400 to 46,800
Option 3 (Hours for Phase 1)	Standalone Search- Proof of Concept Only	3	200 to 400 hrs.	35,000 to 70,000
Option 3 (Hours for Phase 1 and Phase 2)	Standalone Search- Simple Working Site	5	300 to 600 hrs.	52,500 to 105,000
Option 3 (Hours for Phase 1, Phase 2 and Phase 3)	Standalone Search- Full Featured	10	1400 to 2000 hrs.	245,000 to 350,000
Ongoing Maintenance	Search Relevancy Tuning	3	320 hrs. per year (2 weeks each quarter)	80,000 for year

## Recommended Solution

We recommend that ALA move toward a standalone open-source search implementation (option 3). This approach allows ALA to umbrella your many diverse content areas effectively. Along with technology implementation, additional activities around metadata improvement (see the section below on Schema.org) will benefit public search, and establishing a governance strategy and reporting will support content quality and reduce content manager efforts. The recommended implementation will require interaction with many different sites with different security systems to be crawled and exposed in an effective way. Search could bring this to the forefront of your site, increase membership, and strengthen your reputation.

In addition we recommend that after you complete the initial implementation, you stay engaged with the search experts for quarterly tuning, to ensure that the quality of results remain high over

time. At OSC, we provide 4 quarterly tuneups per year to provide this kind of support, with a current cost at 80,000 per year.

Making improvements to ALA search is not a “*big bang*”. It is an incremental series of steps, working to a long-term plan. It is able to be funded and managed in stages, to align with the capacities of the association.

## Summary

As the largest association advocating for libraries and librarianship in the United States, the ALA can set a strong example for search and deliver value to many different interest groups. And you have a novel challenge — ALA is not one large website, but an amalgam of many sites, journals, conferences, and communities. Some of this highly valuable, mission-supporting content is largely hidden. **Quality search is important to meet your user needs and ALA’s goals.**

## Introduction

The American Library Association (ALA) and its Office of Information Technology and Telecommunication Services (ITTS) recognize the importance of search as a vital part of the association’s information service delivery to its members and the public who care about the future of libraries and the value libraries have to transform communities and foster life-long learning.

ITTS is currently planning for both short-term and long-term improvements to the ALA.org website. As part of that planning, ITTS asked OpenSource Connections (OSC) to support them in a rapid analysis to identify options to move forward on improvements to ALA.org and its associated sub-sites’ search capabilities. OSC spent two days with ITTS in December 2015, in order to assess ALA’s current situation and needs and to identify technologies and content considerations to improve search. This report summarizes our findings and includes options and recommendations for ALA’s consideration.

# The Challenges

## Organizational Content Challenges

### Disparate organizations and identities

There are dozens of offices, divisions, round tables, and interest groups within ALA. As a result, there are dozens, perhaps closer to hundreds of individual content sites and collections in a number of different content management systems and social media sites. ***Search is one of the few online unifying elements ALA can offer, thus it is very important to the success of ALA's groups to serve their members and audiences.*** However, ALA has no search system capable of delivering results across all of the sites where content resides.

### Disparate audiences, experience, and needs

Providing easy access to information across all the different areas of their operations—advocacy, education, conferences, memberships, professional collaborations, publications, and more is a challenge to all associations. ALA's situation is compounded by additional challenges based on the professional market that it serves. There are many facets to librarianship and library advocacy and many different audiences that have very particular, specific interests and needs. Website and search analysis reflects this, showing an extremely “long tail” where only ~6% of users arrive via the ALA home page, and only a handful of pages garner more than 1% of the initial starting points for users. ***Search becomes one of the most direct ways for people who know what they want to navigate to it directly, and for people who don't know what they want to see the wide range of available resources.*** However, it is hard for search to be “*all things to all people*” without a sophisticated underlying set of well-designed capabilities.

### The “ALA voice” is diffuse

Advocacy, information policy, and professional leadership development are important aspects of ALA's mission. These focus areas are reflected in many different ways, including publications, conferences, position statements, educational materials, committee and working group documents, membership information, and much more. Beyond ALA as a whole, the divisions also have their own strategies and content priorities. Given the disparate audiences and messages: *How does ALA surface up what is most important to its mission? And how do people (members, the public, policy-makers) find what is most important to them, when they need it?* ***The upcoming ALA website redesign will certainly help simplify and unify the online presence, but it must be accompanied by an effective search capability to get people to the most valuable information, no matter where it resides.***

## Content managers have many other responsibilities—and your members have high expectations.

In most cases, ALA staff and volunteer content creators have many other demands on their time in addition to preparing content for the various ALA websites. When many of ALA.org's primary users are professional catalogers and information specialists, they expect ALA to be at the leading edge, to be an exemplary example. **However, search is driven by the quality of content and structured information/metadata, so these two elements are crucial to site search effectiveness.** But with little time, and many different technical platforms for publishing content, it becomes easy for content managers to skip the step of adding metadata, not realizing the impact that has on users. The site's existing classification schemes and metadata requirements reflect the sensibilities of professional catalogers, not busy part-time content managers.

If ALA wants its message to be heard and the value of its content to be available to its users, then simplification and consistency, supported by sufficiently sophisticated technology are important drivers for both content producers and content users.

## Problems with Search

Let's start with a simple fact: **ALA.org search isn't working effectively for its users, its content providers, or ALA as a whole.** Even with recent changes, it's not possible to "find what I need on page 1" of most search results.

- The diversity of users, their topics of interest, and the organizational services/offerings, make it very difficult to tune a search to identify "the most relevant" specific result for each user.
- In the overall ALA site search, a significant portion of important content is not indexed for search, because it is either spread across such a huge number of disparate sites (which manage content using various different content management systems), or it is caught up inside large non-web documents (PDFs and other types of files) that are challenging to index effectively.
- Conversely, the more content that is indexed and added to search results, the harder it is to find one specific thing—because there will be a lot of similar content competing for attention and priority on the results list.
- There is a tension between the desire for completeness in capturing metadata (particularly topical information about the content) and the time, motivation, and capability of the staff to fully populate the detailed metadata that is currently required.
- Search results don't bring value to the surface, and thus don't drive traffic to revenue-generating and user-supporting content.

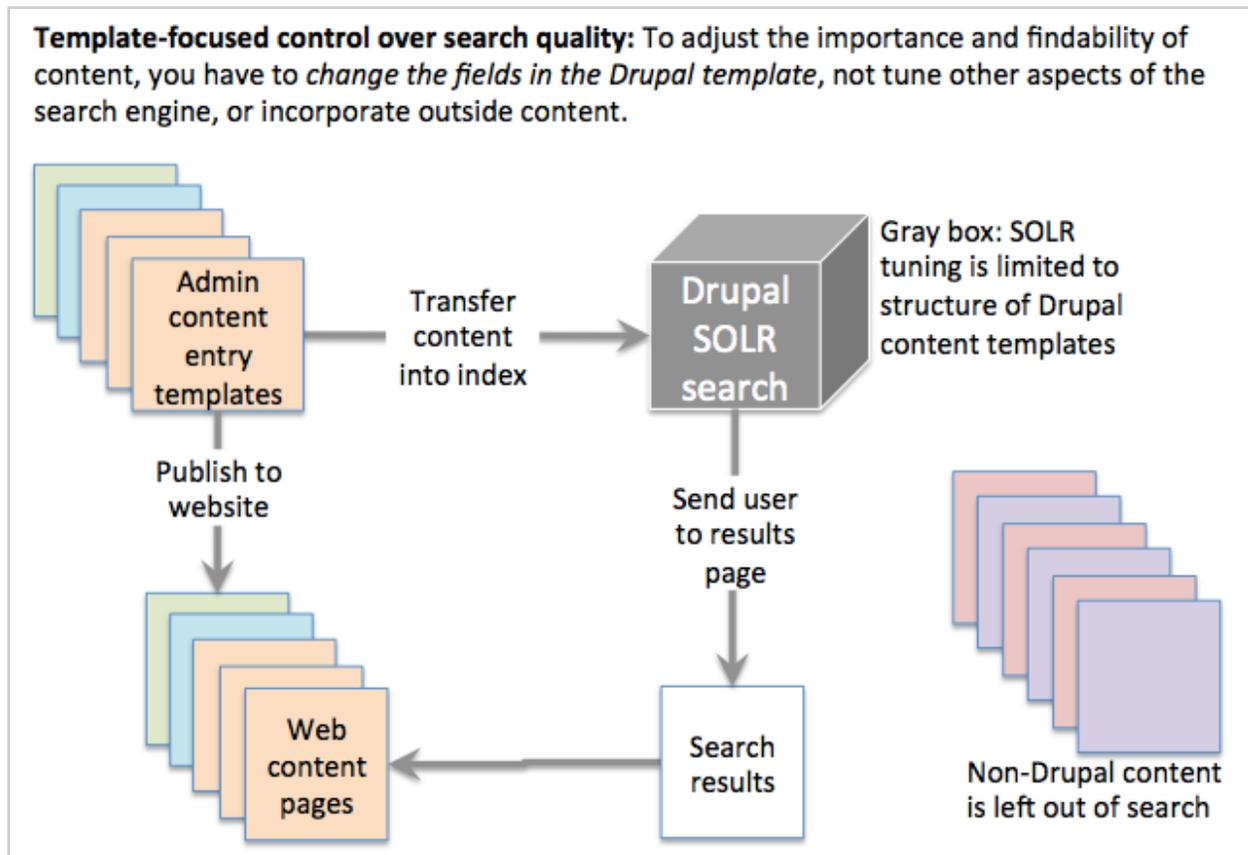
## Problems with the Previous Drupal-based Search

During our analysis, we found two critical problems with the Drupal-based search that has been used by ALA until very recently. The first is that it unduly restricted searching to the content available inside of a single or small set of Drupal websites. A substantial amount of the content available under the ALA umbrella was not searchable. Most importantly, the high value content of conference and conference proceedings sites, journals, and the on-line store are missing. The second issue is that Drupal's web interface for the Solr configuration made it very difficult to properly tune relevance of ALA's documents because it focuses mainly on how the Drupal page templates are built and managed, which in ALA's case is not sufficient for tuning. Only someone with significant experience in tuning Solr could make effective use of that interface—in which case it could make more sense to work directly with Solr, gaining advantages in tuning precision. Given ALA's very large body of diverse content formats, Drupal's Solr plug-in provided tools that seemed simple and useful, but led to very poor relevance. This forced searchers to wade through many results (high recall) in order to locate specific items of interest, assuming that the items were even on the indexed sites, and not elsewhere.

In summary:

- Drupal-based search configuration is not an adequate tool to tune search performance for the content it controls (content within Drupal sites).
- Drupal-based search fails to include a significant portion of ALA's important content (content in other systems, and some types of documents that are posted online).

The following illustration shows the way that Drupal-based search works, and why it is extremely difficult to manage configuration and tuning, along with what is not searched.



## Problems with Google Custom Search

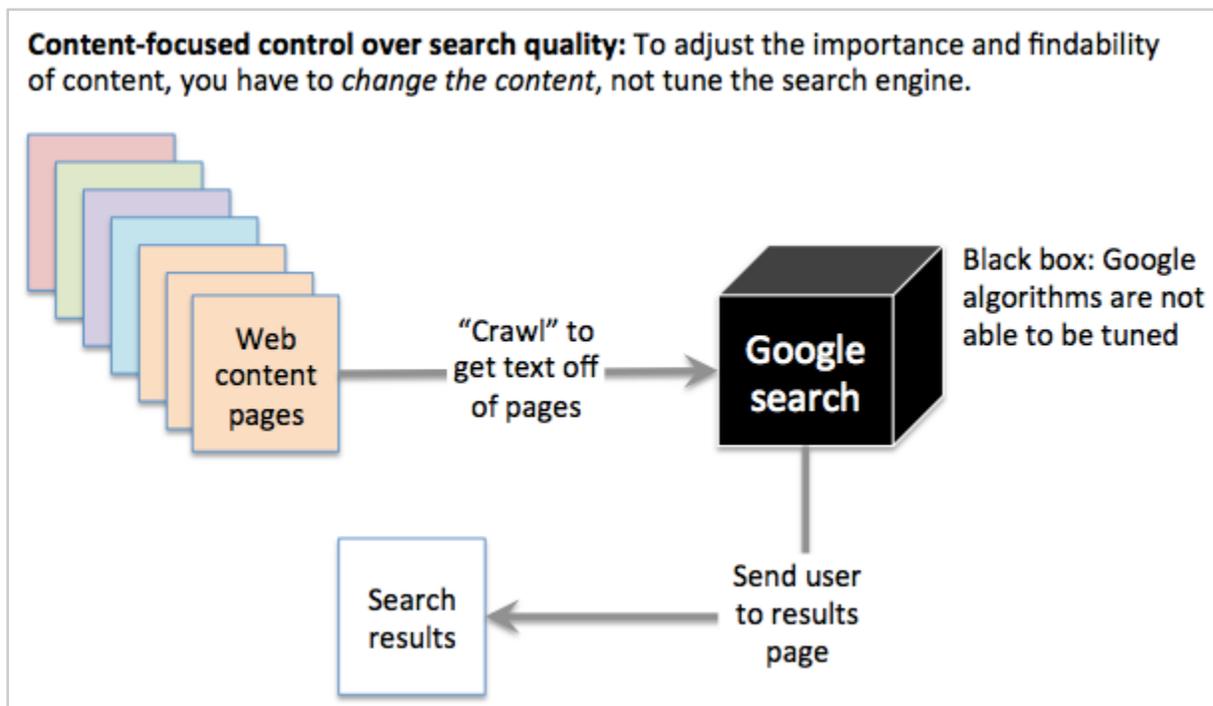
ALA has recently changed the search tool for the majority of their content sites on ALA.org to a cloud-based service: Google Custom Search (GCS). This change removes a primary challenge of many ALA sub-sites on platforms outside of ala.org not being indexed or incorporated in search results, which was experienced with Drupal-based Solr. GCS currently offers up results on ALA.org from any ALA domains that include “*ala.org*” somewhere in their URL. However, there are challenges with GCS as well:

- The quality of results relevance ranking is dependent on the quality of both the content and the metadata in the documents published on the various ALA sites. If that content is not clear, specific, and carrying appropriately structured and complete metadata, the results are poor.
- There is no ability to control the weighting for importance, so certain types of content and important individual items may not be ranked as most important. This can affect content

such as standards, ethics guidelines, bill of rights, annual reports, advocacy communications, etc.

- There is no ability to provide additional filters to help users manage the volume of content returned in results, or to be more precise about what they want to find.
- There is little or no ability to change the format or layout of the search results, which could help users evaluate the results items they receive, and also reinforce branding decisions for ALA and sub-sites.
- There is no ability to provide access to members-only materials.
- Sites like <http://2016.alamidwinter.org/> whose domain names do not include ala.org are still not included in the results.

The following illustration shows the way that GCS works, and why there are no adequate controls to manage the tuning or the presentation of search results.



# Looking Forward: Solution Options

## Overall Attributes of an Effective Solution

1. **Open** - Can work across multiple sites (not just one Drupal instance or sets of Drupal instances) and different content management systems, including WordPress.
2. **Agnostic** - Avoid results ranking advantages based on format (HTML, DOC, PDF). This is another shortfall of Drupal and its relevancy controls.
3. **Configurable** - Allow fine grained control of the relevance of all types of content from all sources: Web pages, documents, and other types of content (e.g. eLearning, media).
4. **Staged** - A solution can be delivered in stages for agile deployment and adjustment.
5. **Service-Oriented** - It should be easy to incorporate a search and results page/component into any website or CMS system.
6. **Distributed** - It must allow for the autonomy of individual divisions, offices and groups to create content, probably on different platforms.

## Benefits of a Complete Approach to Search

Given the above problems, there are a number of things that must be done to improve search. Implementing a search engine for a complex, multi-site environment is not “magic.” It requires:

- Careful technology selection
- User-centered design that considers the role of search in the context of ALA’s web experience as a whole
- Sufficient (but hopefully not over-burdening) metadata and topic management
- Over time, effective governance and quality control of the content that is searched

Also, search is not restricted to ALA’s internal site search. Over 50% of the traffic to ALA’s main website comes via external public search sites (e.g. Google, Bing, Yahoo, etc.). Improvements that ALA staff makes to its content, metadata, and topic tagging will also be used by these external search sites and will boost the availability and ranking of ALA content generally. One aspect of our content management recommendations is to adopt schema.org classification schemes for your web content. Doing this allows you to embed metadata in the modern “*knowledge graph*” linked data format that is used by major search engines that founded schema.org specifically to encourage institutions to standardize metadata formats in content for better searchability. ***Thus, improvements to ALA’s site search should also provide improvement to ALA’s external public search rankings.***

## How Search Supports ALA Goals and Expectations

ALA provides leadership for libraries and the profession of librarianship. Information produced by ALA is a critical foundation to achieve that mission across the country and the world. While “*search*” may appear to be a small part of the online services of the association, it is one of the primary ways that users find online content. Everything that can be done to improve online search capabilities increases the value and authority of the organization.

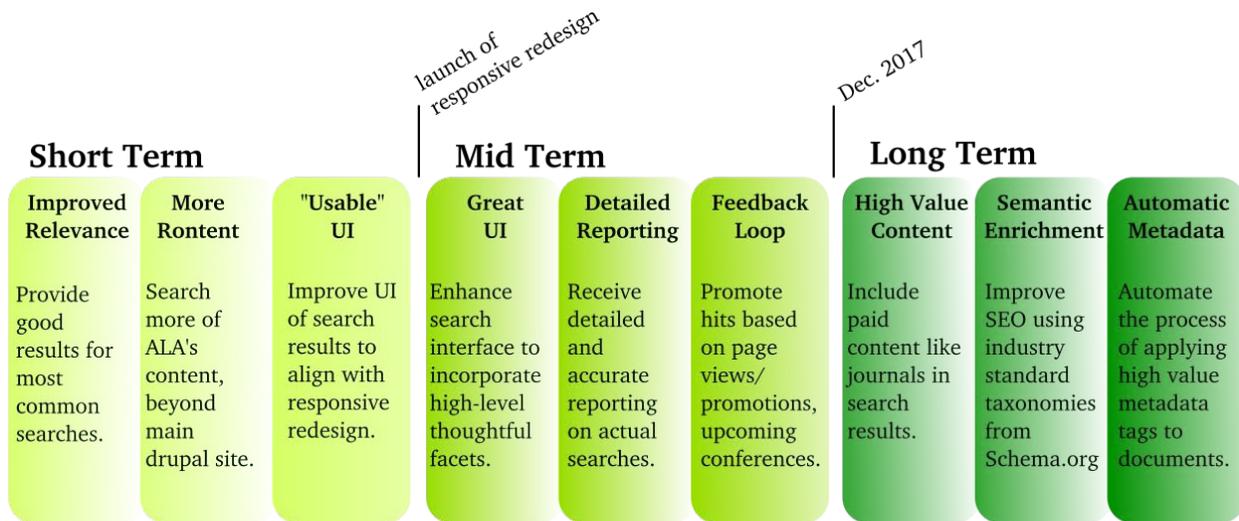
Timely access to relevant, valuable information must be an ALA priority.

- For librarians and other members, their jobs—and their support of others—are based on being able to rapidly access whatever they find most relevant from a wide array of resources. This might include technical information, recommendations, education and certification content for professional development, events and conferences, guidelines, policies, and standards, and community support.
- For teachers and the public, understanding what content is available and valuable for their needs supports advocacy and educational goals.
- For advocates and legislators, it is important to quickly access ALA’s policy positions and evidence for action, and find the most relevant information to reinforce the role of libraries in society.
- For ALA divisions, offices, and round tables, a quality search capability ensures that their resources are available to the widest possible number of people.
- For ALA staff, timely access to information allows them to efficiently support members and users, coordinate information from multiple parts of the organization, and thus support the mission and services of ALA.

Libraries have a critical role to play in our society as they work to transform communities and individuals. Helping people find and use information resources is a vital component of these efforts.

A new ALA.org responsive website redesign will be launched during 2016. A consistent, supportive search experience that aligns with the new design will be critical to its success. This is particularly true if one goal is to improve the mobile experience. In the longer-term, a sustained commitment to quality search will continue to allow ALA to provide a vast array of resources in the simplest, most consistent manner.

## Short-term, Mid-term, and Long-term Capabilities



---  
 "Improved Relevance" and "More Content" are efforts to address the issue of *disparate organizations and identities* as outlined in the introduction.

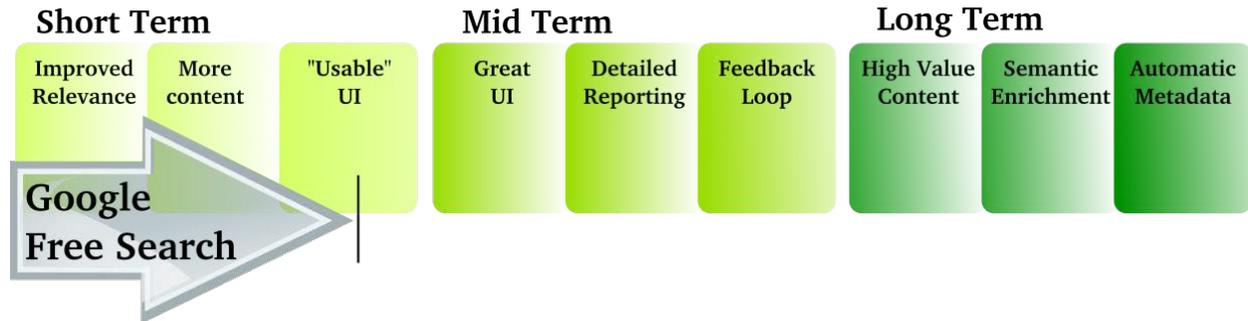
To address "*disparate audiences, experience, and needs*," ALA will need to have more than an attractive search interface. The association will need a "great user interface" that helps direct your users to the content they need. This can be accomplished with faceted searches that align with the top level navigation developed for the new responsive redesign.

**Content managers have many other responsibilities and your members have high expectations.** To address these issues, we want to provide "*Detailed Reporting*" to help your managers understand what people are trying to find and how they consume content. A good "*Feedback Loop*," "*Semantic Enrichment*" and "*Automatic Metadata*" are ways to add critical metadata into the system without demanding more from already busy content creators.

## Technical Options

- **Google search**
- **Paid 3rd party service**
- **Stand-alone service**

### Technical Option 1: Google Search (Free)

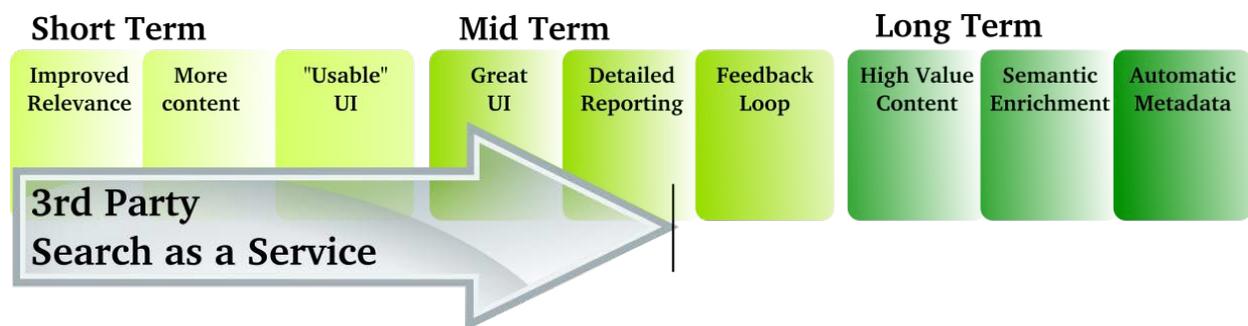


Google Based Search provides an inexpensive short term solution. It is quick to implement, low maintenance, and low cost overall. Further, it takes advantage of any SEO built into the myriad sites that are a part of ALA.org. It will further encourage good SEO throughout the community. It also greatly expands the amount of content that will be searched—anything that falls under the ALA.org domain can be searched in a single location.

Google and all third party services reviewed will unfortunately fail critically and immediately in searching the full extent of ALA’s content offerings, as you are not “one site” but a multitude of sites and services.

Google will not integrate with the new responsive redesign currently being implemented as it cannot be styled significantly. It also cannot be connected with the new top level navigation through faceted search (imagine the ability to restrict results of searches based on the terms you use in top level navigation, which will be familiar to users). Control over the results format is very limited and extensibility is non-existent. Think of it as a slick black box. Not only can we not see inside it, but it does not allow us to connect to it, enhance it, or alter its behavior.

### Technical Option 2: Paid 3rd Party Service



Google provides a paid version of their “*Search as a Service*” offering. It comes with some ability to style results, request immediate indexing (high effort) and a few other tiny bells and whistles. Cost is based on queries executed per year. Expect to pay around 2k to 4k per year based on the 1.4 million queries that were submitted over the last year to ALA via ala.org.

There are many more options for paid services than Google. SwiftType is an excellent example. Companies like SwiftType are based on the Google model of “*Search as a Service*.” They are just as easy to integrate as Google, but often provide a much higher degree of control over search results. For example, SwiftType offers substantially better faceting, custom ranking, synonyms, spell checking, and weights. A huge benefit of SwiftType is real time indexing. Unlike Google, enterprise plans will provide near immediate updates to your index, making content instantly available.

SwiftType leads the pack in creating a very strong positive (albeit generic) user experience. They employ type-ahead logic and respond very quickly—a critical attribute. Many third party services provide a certain amount of detailed reporting and feedback. Though your ability to act on this data to enhance search is very limited, it can be used to guide policy, refine navigation tools, and focus attention on trending topics.

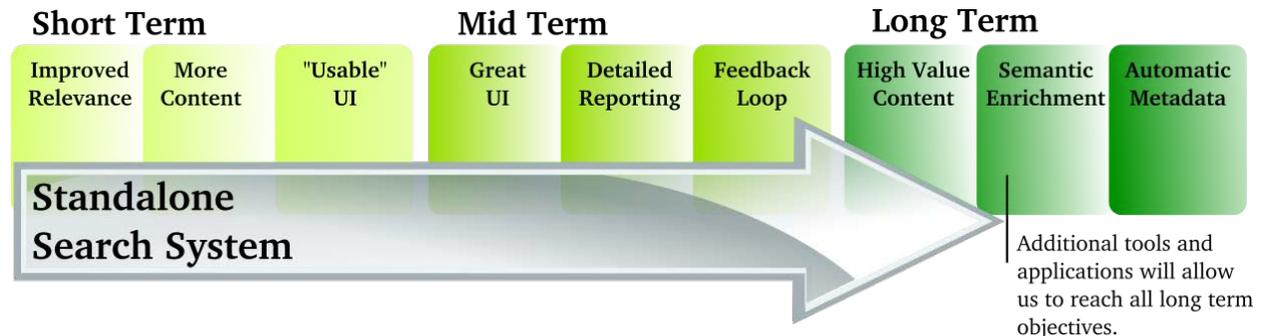
ITTS will need to talk to SwiftType directly to negotiate a financial agreement. Their pricing model is on a site by site basis, and ALA has many sites to index as a part of a single search. The number and extent of ALA’s sites will affect pricing and the feasibility of this approach.

Much like the free Google offering, these third party services are a black box. Once put in place and configured, there is little ITTS can do to further their feature offering. A huge downfall of both Google and SwiftType is that they are targeted for small to medium size websites, not large multi-site associations. So the tools available to the association now and the new features that come in the future will not target ALA’s needs.

In the long term, this system still isn’t indexing all the content. Much of ALA’s revenue generating materials —such as the conference sites, the proceedings and journals, BookList, Connect, JobLIST etc. will not show up at all. They are, in effect, actively hidden, and not helping to promote the value of ALA membership.

Metadata for third party vendors are simplistic and restrictive. Metadata added to a third party system are not portable. ALA could invest thousands of hours tagging content, and then lose that effort when moving to a different system. Complex taxonomies, automated metadata discovery, and integrating with any other tools or systems are non-starters.

### Technical Option 3: Standalone Search



Tools like Solr and Elasticsearch provide a high degree of freedom and customization. They are the only tools that will give ALA a chance to meet the full set of short and mid-term goals. Coupled with tools like Nutch (a web crawler) this is the only option able to fully crawl all the content you have into a single search interface. There is a very high degree of control over relevance ranking, faceting, and synonym usage. You are no longer tightly coupled to a third party vendor whose features or costs are subject to change. The search results interface is fully under your control, which opens up tremendously powerful (and potentially profitable) features down the road.

In the short and mid-term, developing a custom standalone search system will come at a higher cost with greater time investment from internal staff. In its first iteration, the interface may be less polished than some third party search services, though better than Google's free offering. Since ALA already has a Solr installation that can be improved upon, adding Nutch and tuning the results makes for a fine mid-term solution. Not only can users search across the bulk of ALA's content, but ITTS is in a position to gather detailed metrics on search usage across that content, glean a far more comprehensive understanding of your audience for use by Membership, Marketing, and the divisions. Because the search interface is fully in ITTS' control, ITTS staff can deeply integrate a cohesive set of facets that fully realize your efforts based on defining a consistent metadata standard. The interface could not only be a good user experience, it could be a real innovation.

ALA should seriously consider bringing in a team of experts to help properly index the content, tune the results, and create an effective interface for displaying search results. Once a baseline is established, ALA will have an open system that can be easily extended and augmented— so long as the association continues to invest in its development and maintenance.

## Additional Capabilities and Technologies

There are additional technologies to consider for the future as ALA moves forward—to improve search and also the overall use of content on ALA’s websites.

**Schema.org:** As mentioned previously, one of the important actions for ALA to take is to begin embedding simple standardized metadata into its web content pages. Over the past three years, the external public search leaders (Google, Microsoft Bing, Yahoo, and Yandex) have been creating a standard for how content providers can embed metadata into their web pages, which will be read and used by the major search engines to improve result rankings. This simple embedding of structured metadata (particularly categorizations and subject keywords) is increasingly incorporated into major content management systems like Drupal and WordPress, along with common commercial applications. Schema.org coding is also consumed by major search engines, adding a significant degree of control to ITTS for managing results relevance. It is not without some effort—the content needs to have metadata in the first place—but the benefits of applying even simple, high-level metadata to ALA content will pay off in higher search rankings and more exposure,

There are three general activities involved when incorporating schema.org metadata:

- Reviewing the schema.org ontology structure to identify which metadata fields are applicable to ALA content. This activity is likely to take only a few days to assess.
- Identify the implementation requirements for each content management system and website/subsite. This activity will be fairly simple for sites based on Drupal and WordPress, because they have modules available for those platforms. Other platforms, and hand-coded sites (if any), will need to be assessed for available code from suppliers, or the effort to incorporate the metadata structure into web pages.
- Identify the current level of available metadata in ALA content, compared against the selected schema.org fields. Determine the content management editing required to provide any additional metadata for important sites to be found via external search.

**Rich analytics:** Given the diversity of content produced by ALA, the array of audiences, and the large numbers of authors, it is important to know as much as possible about how content is searched and used. This is not just “nice to know” information. It can be applied back into the search algorithms to help tune relevance ranking. It can also be fed back into reports for content authors to improve (and in some cases automate) how keywords are populated.

**Thesauri and alternative dictionaries:** These are commonly able to be incorporated into the Solr or Elasticsearch engines. One thing for ALA to consider is how to incorporate external

vocabularies to extend and refine the metadata in the association's content. Many useful, simple vocabularies are now becoming available as linked data sets for use when improving search.

**Machine-supported concept extraction and tagging:** The quality of “*automated tag suggestions*” has improved dramatically in the past few years and has also become more affordable. The best practices that are emerging use the technologies as support for authors—they do not automatically add keywords to content, but suggest relevant keywords that authors review and confirm before publishing. This is known as “human in the loop” concept analysis. This saves authors significant time creating metadata and improves consistency. There are a growing number of companies who have these tools and services, often in combination with taxonomy development services—some of the better known ones include Access Innovations (the Data Harmony tools), Synaptica, SmartLogic, and Temis, and there are others. Their prices will depend on many factors, so this is not included in estimates. And while there are some emerging open source and cloud service capabilities, they have not necessarily been as domain-relevant, robust, or supported by consulting services. This area requires further research as requirements are clarified.

**Deeper text analysis and summarization:** The machine learning community has been working on an array of tools that would allow ALA to “*mine*” the relevance criteria and terms from larger content (PDFs, learning content, slide decks, and other types of large documents) to better control how they are managed by search, as well as improve how they are presented in the results interface.

**Multilingual vocabularies and designs:** Particularly in public libraries in the United States, it has become important to accommodate multiple languages (primarily, but not only, Spanish). Search engines and content applications are increasingly going to be incorporating tools that help content providers support this need. These should be regularly reviewed by ALA for inclusion in enterprise search.

## Design and Content Requirements

Alongside the technical requirements to improve search, it is important for ALA to focus on users' needs and how search improvements are incorporated into the designs of the association's various sites. ALA ITTS has a strong team that provides internal user-centered analysis and design. The following are some things that team must consider. In some cases, ITTS can be supported by external partners who are able to provide specialized skills or more focused resources.

**Content strategy:** Ultimately, search implementations cannot be successful without clear, relevant content and somewhat complete metadata. Decisions about content creation, format, categorization, and management over time are widely distributed throughout hundreds of content managers at ALA. In parallel with the focus on search technology, we recommend inventorying existing content and identifying improvements to be made. This should focus on user needs and the optimization of search/related links throughout the site, but also on reducing the burden on content managers to “*get it right*” and maintain content effectively. One early task could be to review the last couple years of site traffic analytics to identify where some content is not being accessed by users. Then decisions can be made about priorities, and also establishing an approach for regular archiving of content.

**Prepare a content governance plan and strategy:** Search improvements are a long-term investment. As noted above, they rely very heavily on the quality of content. A governance strategy is essential for ALA to support its very diverse content creation approach. A plan for training, supporting, and evaluating content managers will help the organization manage the search improvement process and ongoing maintenance. A governance strategy also informs the types of reporting and analytics that are required on your websites, to understand how content is found and used in order to provide a regular feedback loop to content managers.

**Analyze the wide range of user search needs:** One thing that was clear in our initial analysis was that ALA search has an extremely “*long tail*” — a vast majority of content that gets a small percentage (less than 1-2%) of overall web traffic, and what appears to be a huge variance in what people are seeking when they search. This makes it much harder to make decisions about how to tune the search engine for ranking as well as how to lay out the results page. A focused analysis of user needs will provide significant information to support search design. It is not yet clear how much of that long tail of content access is because higher-value items are not sufficiently easy to find, or because there is wide variation in user needs for specialized content. The analysis that supports content strategy will begin to provide insights into these issues. This analysis will also inform ALA’s archiving strategy.

**Search results design (ALA.org):** Search results pages are a particular design specialty, especially when they involve federated searches across widely heterogeneous content. In many ways the design will conform to common “norms” in the industry, but in other ways it needs to have particular designs prepared for its unique requirements.

**Search results design (other ALA sites and pages):** ALA.org is not the only place that search will be affected. To create consistency and a common “*feel*” to ALA sites, we recommend looking at all the other ALA properties and their search interfaces. This does not mean all sites

need to look the same or have exactly the same search—it means that there needs to be a “family resemblance” and functional consistency, or it risks creating user confusion.

**Additional search “components” for other site pages:** There are many places where it will be valuable to launch a new search for some particular part of the ALA information space (e.g. “*search within*” a specific area,” or “*find more like this*” across all of ALA). There will also be many places where it will be valuable to present “micro results” such as related links associated with an article. These capabilities can be designed as components that can be embedded in different sub-sites and then managed by background search services.

**Routine review during design:** As it is likely that the ALA ITTS team will be carrying out the search design, we recommend that a regular review process be established with an external search specialist to provide feedback on design decisions. The specialist can respond to any questions or concerns about aspects of incorporating search throughout the various ALA sites. The broadly federated search approach required for ALA and the nuances of how users control the results will require a lot of decisions that relate to the underlying technical capabilities and implementation approach.

**Usability testing and regular user feedback:** It is important during the design and development process to test the designs (and underlying assumptions) with users. This is particularly important for search results designs because there is a high degree of cognitive interaction with the user’s expectations and mental models of what they are seeking and how they search. So usability testing is important, and requires interaction with real content. Usability testing also needs to be done with disabled users, in line with ALA’s accessibility initiative, because search is important as a means of very direct, efficient access to content.

## Recommendations

This section is meant to provide an order of magnitude of effort comparing the options that were identified during the assessment are the basis for the cost estimates. Specifics on the recommendations follow.

Option from Assessment Report	Option Description	Magnitude of Effort	Range of Hours	Range of Cost (in USD)
Option 1	Google Search	1	60 to 180 hrs	11,700 to 35,100

<b>Option 2</b>	<b>Paid 3rd Party Service-Hosted Search (Swifttype)</b>	3	120 to 240 hrs	23,400 to 46,800
<b>Option 3 (Hours for Phase 1)</b>	<b>Standalone Search- Proof of Concept Only</b>	3	200 to 400 hrs	35,000 to 70,000
<b>Option 3 (Hours for Phase 1 and Phase 2)</b>	<b>Standalone Search- Simple Working Site</b>	5	300 to 600 hrs	52,500 to 105,000
<b>Option 3 (Hours for Phase 1, Phase 2 and Phase 3)</b>	<b>Standalone Search- Full Featured</b>	10	1400 to 2000 hrs	245,000 to 350,000
<b>Ongoing Maintenance</b>	<b>Search Relevancy Tuning</b>	3	320 hrs per year (2 weeks each quarter)	80,000 for year

If you have little or no budget, and search is not a priority for your organization, then we recommend staying with Google's free option to meet the basic expectations of most users.

If search is a low level priority that ALA can commit \$20,000 to \$30,000 to over the next three years, then ALA should consider a third party service as described in the second option above. This would require a very low level of commitment from staff, while providing ALA with a professional system that could help the association target specific categories of users.

As the largest association serving librarians and libraries in the United States, the ALA should be setting the standard for search. However, ALA has a novel problem—not one large website, but an amalgam of many sites, journals, and conference proceedings. Some of this content is highly valuable and largely hidden. This will require interaction with many different sites with different security systems to be crawled and exposed in an effective way. Search could bring this kind of content to the forefront of your site, increase membership, and strengthen your reputation. To move forward would require an investment of at least \$200,000 for an initial deployment, along with a serious long term commitment to the project's success through follow on work and adequate staff to support the increase in infrastructure.

## The Potential Role of Partner Organizations / Companies

ALA is not the only organization that is carefully reviewing its search capabilities. Many organizations are reviewing the same technologies and design considerations. We recommend that ALA consider where it can find partnerships with other non-competing associations to collaborate on tool research, design, and implementation to spread the costs and increase the learning potential. In addition, there are a number of small technology companies working on new, valuable capabilities for federated content search that may be open to partnerships and discounted pilot implementations in order to get more "real world" case studies and feedback.

## Consultant Biographies

### **Daniel Funk, Senior Search Engineer, OpenSource Connections**

Daniel Funk is a Senior Software Engineer with extensive experience in both enterprise and academic search systems. As a part of larger teams, Daniel has created the next generation search platform for Patent Examiners at the US Patent and Trademark Office, developed a new e-commerce system for Under Armour, and led development efforts on research projects out of the University of Virginia.

Previously, Dan was a lead developer and software development manager for RosettaStone and a core contributor to Blacklight, an open source search system employed for library cataloging by universities all over the world (<http://projectblacklight.org/>).

Specialties: Java, Scala, Solr, Lucene

### **Duane Degler, Search Strategist & Principal, Design for Context**

Duane Degler specializes in the design of sophisticated interactive applications and search experiences, with an eye toward making rich data resources usable and relevant. He has led web, mobile and software projects for commercial and government clients in the United States and Europe. Since 2002, he has focused on the unique potential in user experience and design for linked data and semantically-enabled applications.

Duane presents on strategy and emerging trends in design, content, and technology. He is a regular speaker for professional groups supporting associations and non-profits and at conferences focused on web/data technology (KM World, SmartData, Taxonomy Bootcamp, SemTech, ISWC), web/mobile and content design (UXPA, IA Summit, IxDA, EdUI), and more domain-specific conferences and workshops in cultural heritage and healthcare. He authored the “Dynamic IA” chapter in the book *Reframing Information Architecture* (2014). For more information, presentations and publications: <http://www.designforcontext.com/about-us/duane-degler>

Duane Degler and Design for Context have partnered with OpenSource Connections on many search and website projects over the past five years, as their combined design and development capabilities deliver innovation to their mutual clients.