DISCOVERY SERVICES:
A WHITE PAPER FOR THE TEXAS STATE LIBRARY & ARCHIVES COMMISSION

BY
ARTA KABASHI, CHRISTINE PETERSON, AND TIM PRATHER
AMIGOS LIBRARY SERVICES

AUGUST 2014

Funded by the U.S. Institute of Museum and Library Services through a grant to the Texas State Library and Archives Commission
EXECUTIVE SUMMARY

Discussions among libraries that have recently implemented discovery services are likely to result in agreement that implementation was challenging. However, once implemented, librarians are generally happy with their decisions to offer discovery services to their patrons. Based on librarian experiences of both the challenges and rewards of implementing a discovery service, the Texas State Library and Archives Commission (TSLAC) contracted with Amigos Library Services to write a white paper that would include basic information concerning discovery services, as well as an overview of the major discovery vendors. Below is an overview of information contained in this document.

DEFINITIONS

From NISO: These services use an aggregated central index to enable searching across a wide range of library related resources – both licensed and free – from multiple providers. They also offer more sophisticated capabilities and faster performance than those provided by systems relying on federated search technologies.

For patrons: Google-like search experience for all library resources.

For librarians: A web-scale, index-based search service that includes local and remote library resources, including full-text article-level content as well as delivery of physical resources.

CURRENT STATE OF DISCOVERY SERVICES

Basic functionality for discovery services is currently available. Additional work is necessary that will allow this technology to work to its potential: best practices, communication between the discovery service and the content provider, and stability.

ADVANTAGES

- Leverages a library’s existing resources
- Single interface for searching multiple resources
- Clear starting point for research
- Search results are more specific to a patron’s needs than using a search engine
- Results are displayed more quickly than in federated searches
- Allows patrons to see and evaluate what is available immediately as well as those that will take more time

DISADVANTAGES

- Implementation costs for a discovery service include more than the cost of the service. Other costs include staff time to map data elements of the databases,
to understand how the service defines terms like “relevancy” and to determine how to teach and market it.

- Patrons have an expectation that everything is available when using a single search box, but that may not be true.
- Integrating your ILS into a discovery system will take time.

**Limitations**

- Completeness. Not all resources work well in a discovery service.
- Relevancy ranking. It is difficult to perfect relevancy searching when the metadata is coming from many disparate sources.
- Speed. How quickly results display is dependent on many things, and sometimes performance is slower than anticipated.

**Areas of a Library Most Impacted by a Discovery Service**

- Staff who have responsibility for the library’s website
- Librarians who deal with e-resources
- Reference staff
- Librarians who provide bibliographic instruction
- Interlibrary loan librarians

**What to Look for in a Discovery Service**

- Usage statistic functionality
- How relevancy is determined
- What is included in the central index
- How the discovery layer works

**Best Practices for Implementing a Discovery Service**

Steps to implementing a discovery service should include:

- Identify target audience
- Build a central index that includes databases of interest to your target audience(s)
- Set up the authentication process
- Customize the service
- Design the default presentation of the search box
- Refine search and retrieval options
- Test usability
- Enhance the service by integrating with other services such as Blackboard, LibGuides, RefWorks and Zotero
- Provide instruction and documentation of the service
EVALUATING A DISCOVERY SERVICE

- Work with the selected vendor and key project stakeholders to identify and troubleshoot the efficiency of the new system
- Recognize how your new discovery system covers resources and indexes them, what you can learn from usage statistics and relevancy rankings and how your selected vendor interacts with you on an ongoing basis
- Understand the contractual agreements and the type of support you should receive
- Measure resource usage against what the vendor advertises
- Utilizing usage statistics, determine which library resources are being found and which are not
- Determine how your discovery vendor determines relevance
- Determine how you will test to evaluate the discovery service
- Be aware of new developments or enhancements to your system

APPENDIX A: MAJOR VENDORS FOR DISCOVERY SERVICES

- BiblioCommons
- Blacklight
- EBSCO Discovery Services
- Ex Libris Primo
- Innovative Encore
- OCLC WorldCat Discovery
- ProQuest Summon Service
- VuFind

APPENDIX B: WHAT TO LOOK FOR IN A DISCOVERY SERVICE

APPENDIX C: EVALUATING YOUR DISCOVERY SERVICE

APPENDIX D: BIBLIOGRAPHY
Since the advent of Google’s single search box, libraries have wanted to provide a similar experience. Unfortunately, searching across web-based resources is much different than searching across the variety of resources libraries provide. Even so, library discovery service vendors strive to do just that.

This paper attempts to provide an overview of discovery services, including their advantages, disadvantages, limitations and best practices. In addition, a synopsis of the major discovery vendors is provided in Appendix A.

OVERVIEW

Discovery services started appearing in 2009. They were a direct result of three changes in library and technical environments:

1. Less than satisfactory results using federated searching
2. More availability of broadband Internet service for libraries
3. Proliferation of databases each with their own searching requirements

DEFINITIONS:

Unfortunately, there is not a single definition for “discovery service,” and this causes confusion. The NISO Open Discovery Initiative includes a definition which may begin to provide some clarity:

These services use an aggregated central index to enable searching across a wide range of library related resources—both licensed and free—from multiple providers. They also offer more sophisticated capabilities and faster performance than those provided by systems relying on federated search technologies.

(Open Discovery Initiative 2014)

A simplified set of definitions will do to introduce the concept of discovery to persons unfamiliar with discovery services:

For patrons: Google-like search experience for all library resources.

For librarians: A web-scale, index-based search service that includes local and remote library resources, including full-text article-level content as well as delivery of physical resources.

A discovery service should include the ability to search as many library resources as possible, making it as easy as possible to identify and retrieve relevant material. This is a change for many librarians who are used to thinking of the local collections as separate from those that are not in-house. This type of service breaks down the silos and makes this a unified experience, allowing patrons to see everything a library has to offer at once.
CURRENT STATE OF DISCOVERY SERVICES

The basics of discovery as defined above are available now. However, in order to make this technology work to its potential, a few things are still needed:

- **Best practices.** In June 2014, NISO released *Open Discovery Initiative: Promoting Transparency in Discovery*, NISO RP-19-2014. This is a NISO Recommended Practice document presenting best practices for content providers and for discovery service providers. As these best practices are adopted, it will be easier for librarians to compare services, as well as identify those that would be the most beneficial for their patrons.

- **Communication between the discovery service and the content provider.** Configuring content for a specific discovery service is not an easy task and requires content and discovery vendors to work together. This can be difficult as sometimes the vendors are rivals in this or other areas.

- **Stability.** Librarians are starting to see this functionality as one that will be important to their patrons; however, until the first two items are dealt with, many will stay on the sideline, not having the resources to make the service work well.

ADVANTAGES

Why do librarians consider a discovery service? Primarily, because they want to leverage their existing resources. Libraries spend money on collections, databases, and other materials so they can be used. When they are difficult to find or use, the organization is not getting as much out of that investment as it could be.

Discovery services provide a single interface for searching multiple resources, e.g., integrated library systems (ILS) and databases. It provides a clear starting place for research. Learning a single interface is much easier for both librarians and patrons than learning a separate interface for each resource. Because it is easier for librarians to instruct patrons in its use, staff have more time for reference interviews, discussion and evaluation of patrons' results. It is also more likely that the single interface will be used more effectively, as it will be used more frequently.

Discovery services present a more effective alternative to both Internet search engines and federated search tools. The set of resources presented to the patron by a discovery service should be more specific to his/her needs than those found using an Internet search engine, as library resources are vetted by the library. The discovery service places the full resources of the library in front of the patron.

Discovery services, which search local indexes, return results more quickly than federated tools, which search databases remotely. Patrons using discovery services do not have to contact a remote server until they want to see the resource itself.
Because discovery services allow library patrons to simultaneously search full-text e-content and library catalogs, patrons can see and evaluate resources that are available to them immediately as well as those that will take more time to access or acquire.

**DISADVANTAGES**

The annual cost of a discovery service is assumed; however, there are other, more hidden costs. Some of these include staff time to:

1. Map data elements of the databases
2. Understand how the service defines terms like “relevancy”
3. Develop effective teaching and marketing strategies

Patrons have an expectation that everything is available when using a single search box. There are some databases or types of data that do not work well or do not make sense to include in a discovery environment. Examples are included in the section “Limitations of a Discovery Service” found below. Ultimately, there may be specific resources that must be searched outside the discovery service.

Integrating records from your Integrated Library System (ILS) into a discovery service may be challenging. Discovery services were created primarily to bring together full-text and other electronic data. Records from an ILS are quite different and may require more time to tweak. Crosswalks are usually available, but depending on your data, they may require some fine tuning as well.

**LIMITATIONS OF A DISCOVERY SERVICE**

There are still issues with the available discovery services. Some of these limitations include:

- Completeness. Not all resources work well in a discovery service.
  - Abstracting and indexing services often come from disciplines with specific vocabularies that provide great value to the discovery process. (Breeding 2014, January 14) These sources are best searched using their native interfaces. Some examples are WestLaw and LexisNexis for lawyers and CINAHL for nurses. These are very precise professions that require precise searching capabilities.
  - Other databases may be difficult to integrate into a discovery service because of their unique content. Two examples include HeritageQuest (genealogical resources) and Learning Express (career resources, tests, and tutorials).
  - Because of the difference in the type and breadth of metadata, some would argue that ILS data should not be part of a discovery service. MARC records provide subjects and sometimes tables of contents and summaries. Metadata is much deeper for electronic materials which can include abstracts, thesaurus terms, descriptions, and full text.
Other resources may not be part of the discovery vendor’s index because the
discovery vendor has not yet implemented the database, or because the
database vendor will not allow it to be implemented. If a resource is not part
of a discovery vendor’s index, it may be accessible via Z39.50, e.g., a
federated search.

- **Relevancy Ranking.** It is difficult to perfect relevancy searching when the
metadata is coming from many disparate sources.
  - The fullness of the metadata provided in the records being searched affects
    relevancy. The more full-text databases or enhanced metadata a library
    obtains, the better the results for patrons.
  - Known item searching is usually weak in discovery services. Items that
    should be at the top of a result list often are not.

- **Speed.** Most, if not all, discovery services are cloud-based. How quickly the
results display depends primarily upon the speed of the Internet connection
from beginning to end. Reasons that a discovery service performance may be
slower than anticipated include:
  - The server(s) used by the discovery vendor are undersized or inundated by
    the number of patrons.
  - The bandwidth on the vendor’s side or on the library’s side is too low or
    inundated by the number of patrons.
  - Results from Z39.50 resources may be slower to display than those from a
    vendor’s index.
  - Search results display quickly, but the full-text may be slower to display
    because the file is being retrieved from the original server, not the discovery
    vendor’s.
  - The full-text being displayed could be a very large file and might take time to
download and display.

**Areas of the Library Most Impacted by Implementing a Discovery Service**

- Library staff who have responsibility for the library’s website will be involved in
integrating the discovery service into the website.
- Librarians who deal with e-resources will be involved in culling through
databases provided by the discovery vendors in the central index. In addition,
they may be needed when adding other resources the library owns or subscribes
to.
- Reference staff may be involved in configuration discussions.
- Librarians who provide bibliographic instruction will update their materials and
change the way they teach, motivating patrons to use a single search box.
Instead of teaching multiple interfaces, librarians can spend more time
discussing the evaluation of resources.
- Interlibrary loan librarians should consider their workflow and processes, as
implementing a discovery service may increase use of the collection and, if
enabled in the discovery service, of the ILL service as well.
WHAT TO LOOK FOR IN A DISCOVERY SERVICE
A checklist version of this section can be found in Appendix B.

- Usage Statistics Functionality. Usage statistics should minimally include the total number of searches, result clicks, total number of click-throughs, total number of searches per month, total number of unique visitors per month, total number of click-throughs per month, top 500 search queries for the last period, and the top 100 referring URLs to the discovery service for the last period. (Open Discovery Initiative, p.27-29)

- Relevancy Ranking Practices.
  - How relevancy is determined and if it can be modified by the library.
  - Ranking of search results should be objective. This is a particular concern if the vendor provides both content and a discovery service.
  - If specific providers or types of documents can be privileged; that is, can certain providers or types of materials display before the rest of the results? Options could include displaying the library’s holdings or full-text materials first.
  - Does the relevancy ranking provide results helpful for your patrons?

- Central index.
  - Which databases are available in the central index?
  - What is the quality of the metadata?
  - Does it include the types of materials you need, e.g., full-text, citations, journal backfiles?
  - Is the full-text searchable?

- Discovery layer.
  - Does it include advanced searching options, facets, and limiters? Are they easy to understand and use?
  - Does it include end-user features helpful for your patrons, e.g., lists, tagging, citation export, and social media integration?
  - Can you customize the look and feel or branding of the website? Are widgets and APIs available?
  - Can results be enhanced with cover art, recommendation engines, or other external information?
  - How usable is the site for patrons?

(Best Practices for Implementing a Discovery Service)

When implementing a discovery service, most libraries will move through the following steps:

1. Identifying target audience(s).
2. Building central index. The central index should include databases of interest to your target audience(s). Not all databases provided by the discovery vendor need be included in your library’s central index.
3. Authenticating patrons and resources. Not all discovery services work with all authentication options. Libraries may not have a choice in the type of authentication they must use. Setting up authentication can be difficult and time-consuming.

4. Customizing the service, including branding. Part of customizing the service includes mapping data. Because mapping the ILS is often a difficult task, be sure to include those that understand the catalog and how its fields have been used over time in this discussion.

5. Designing default presentation of search box. Lowin, Sierra, and Boyer (2013) suggested a number of items to consider when providing a single search box, as is done in discovery:
   - Think carefully about how you present this search box in context with your other services. In particular, consider how to deal with services that overlap in functionality, e.g., discovery service, ILS, specialized databases.
   - Will this be a primary or a secondary search tool? A discovery search box which patrons use, assuming they are searching the website or just the ILS, can become an obstacle. Use multiple search boxes or tabbed search boxes to direct patrons to appropriate search tools.
   - Search results from a discovery search must not only provide results, but also differentiate among the types or formats of resources.
   - A discovery search box should be centrally located and given increased screen real estate. The search box should use tabs rather than drop-down menus.

6. Refining search and retrieval options. Make decisions such as:
   - Will you display your ILS records first, before other resources in search result sets? This would allow patrons to see what is held in your collections before looking at other databases.
   - How will you take patrons seamlessly from search to fulfillment? When thinking of discovery services, most people consider the viewing or downloading of full-text as fulfillment. However, the service may extend to any material available through the discovery service, even materials available at other libraries.

7. Testing usability.

8. Enhancing the tool.
   - Make search boxes portable so they can be presented within other services like Blackboard or LibGuides.
   - Provide instruction. Although using a single search box may seem simple to use, instruction and documentation are needed for patrons to understand the scope of the search results and the access options for different materials.
   - Allow the integration of popular bibliography management tools like RefWords and Zotero.
   - Create subject guides to supplement the discovery service. Particularly for research that may want highly specific information, the discovery service
may provide results that are too broad. Providing subject guides can aid those patrons in finding the resources they need.

(Thompson 2013)

Do not underestimate the time necessary to customize the service (#4) and refine the search options (#6).

**EVALUATING YOUR DISCOVERY SERVICE**

A checklist version of this section can be found in Appendix C.

Once the institution has made the decision to implement a discovery service, it is important to work with the selected vendor and key project stakeholder to identify and troubleshoot the efficiency of the new system. This is the time to evaluate decisions made during the implementation process and utilize patron feedback to inform and ameliorate future workflows. It is important to analyze the goals of implementing a discovery service and determine how they align in practice. Did you want to give primacy to your local collections, have your patrons discover more relevant material, or make the user experience with library resources easier?

Evaluation also includes recognizing how your new discovery system covers resources and indexes them, what you can learn from usage statistics and relevancy rankings, and how your selected vendor interacts with you on an on-going basis.

**VENDOR COMPANY**

Understand the contractual agreements between you and the vendor. Understand the type of support you will receive from the vendor as part of your agreement and begin to evaluate that experience. Areas to focus on include company stability, quality of their staff, your experience during support interactions (outside the implementation process), quality of help they make available, and the process through which they handle conflict. Be aware of how they handle system updates and technical support.

**RESOURCE COVERAGE AND INDEXING**

Since there are currently no standardized tools to adequately measure how much of subscription content is covered by a given central index, it becomes imperative to measure resource usage against what the vendor advertises. It is important to work alongside your vendor to understand what items are discoverable and perform searches that cover full text, subject headings, and abstracts. In “Paths of Discovery,” Asher, Duke and Wilson, discovered that students were unable to evaluate sources on their own and fully depended on default search settings. This is all the more reason to work with your discovery vendor to establish usable search default settings for your institution. (Asher et al 2012)

**USAGE STATISTICS**

All the major discovery vendors provide usage statistics to their subscribing institutions. It is critical to understand through this data which portion of library resources your patrons are finding through the discovery service, and what type of information they are accessing through other resources such as Google Scholar.
RELEVANCE RANKINGS

Relevance ranking is another critical test of your discovery service. Librarians need to know which resources are rising to the top of their search results for given searches. Many vendors will not share their search algorithm with their subscribing institutions. As such, it is very important to run searches to discover if the vendor’s own data will appear at the top of the search results first or a combination of sources. EBSCO is one of the discovery vendors that provides detail on their website regarding their method for ranking search results.

TESTING

Once you have established criteria to evaluate the discovery service, you must identify the resources and tools through which you will gather testing data. Some institutions rely solely on quantitative or qualitative research methodologies; others use mixed methods. Whichever you select, you must adhere to proper research protocols. It is important to build your network of resources, e.g., other colleagues who are using the same system, and solicit their advice when building test scenarios. Maintain a working relationship with your subscription vendor and utilize their help in connecting you to other sources in your region.

Apart from relying on your colleague network, begin to build your own scenarios based on patron feedback. Build survey forms that display after patron search sessions, and follow-up with patrons via telephone or face-to-face interviews to capture the user experience. Rely on focus group feedback to determine enhancements to the services.

One test scenario might be to evaluate resource discovery with and without the use of subject guides. Divide your focus group into two teams, with one team relying only on default discovery settings, and the other team adding subject guides. Note the differences and compose your report.

TRENDS

Once your testing has been completed in-house, be aware of new developments or enhancements to your system. Begin to collect information from the vendor, e.g., technical information, case studies from other institutions with the same patron demographic. Attend seminars, conferences or other on-going focus group meetings; and participate in focus groups that seek to improve your system’s operations.

CONCLUSION

Discovery services continue to undergo changes, working toward Google-like simplicity, but with library-specific functionality. Activities like NISO’s Open Discovery Initiative (ODI) and on-going discussions via email discussion lists, webinars and conferences serve to increase awareness of this type of service, as well as provide a platform for sharing experiences. In addition, the ODI “. . . aims to facilitate increased transparency in the content coverage of index-based discovery services and to recommend consistent methods of content exchange or other mechanism.” As vendors
re-tool their services to comply with the ODI, it will become easier to compare and evaluate discovery services.
APPENDIX A

MAJOR VENDORS FOR DISCOVERY SERVICES

Vendors chosen for this overview had to meet these requirements:

- Can be used with a variety of integrated library systems
- Either a relatively new discovery service or one that is well-known in Texas

VENDORS

BiblioCommons
   http://www.bibliocommons.com/

Blacklight
   http://projectblacklight.org/

EBSCO Discovery Services
   http://www.ebscohost.com/discovery

Ex Libris Primo
   http://www.exlibrisgroup.com/category/PrimoOverview

Innovative Encore
   http://www.iii.com/products/encore

OCLC WorldCat Discovery
   http://oclc.org/services/discovery.en.html

ProQuest Summon Service
   http://www.proquest.com/products-services/The-Summon-Service.html

VuFind
   http://vufind.org/
BIBLIOCOMMONS

BiblioCommons was founded in 2006 through a project that studied “the emerging technologies that teens were using to engage with popular culture” and how they “might be co-opted to establish a social context for the sometimes isolating activity of reading.” (BiblioCommons 2013).

This discovery service focuses exclusively on public libraries, their public catalogs and typical catalog functionality. Their research has found that public library patrons have difficulty navigating result sets that include both catalog and database materials. As a result, they do not intermingle these results, but separate them in one of three ways:

- Catalog results from a keyword search would display in the main body of the web page. A left column titled “Explore Further” would display results from online or e-book resources.
- When using BiblioCommons to integrate the catalog and a library website, it creates a “Research” or “Databases” tab in the main navigation. This tab could provide access to specific topics or databases.
- Patrons can choose to search the catalog, the website, or databases/articles by using radio buttons under a simple search box.

Because catalog and database results are separated, federated databases can be included as separate options in any of the three options above. BiblioCommons can integrate with databases and digital collections that support SRU and/or either Dublin Core or MARCXML record schemas.

Authentication options include username or barcode and PIN.

The basic functionality of this system is that of an ILS, therefore other types of data, like journal articles, will require mapping. Results, whether citation or full-text, from other databases will display within the original interface, e.g., articles from EBSCOhost will display within their website. For e-books, they currently have a browser-based platform for reading, in beta (BiblioDigital).

Their central index includes the data only within the library’s ILS. All other resources must be purchased or subscribed to by the library.

Minimally, BiblioCommons works with the following ILS systems: Symphony, Polaris, Horizon, Sierra, Millennium, VTLS, Carl-X, and Evergreen.

BiblioCommons’ relevancy ranking is the result of ongoing keyword analysis, as well as incorporating various circulation metrics to automatically adjust to an individual library’s holdings and circulation patterns. Relevance criteria include:

- Overall circulation metrics and holdings across all BiblioCommons libraries
- Ongoing keyword analysis of patron search behavior
- Individual library’s holdings and circulation patterns, ensuring that what is relevant to the community (based on metrics such as the number of titles, number of holds) is reflected in the search results
Library staff can run reports on BiblioCommons activity, including:

- Type and quantity of user-generated content (reviews, ratings, lists, etc.)
- Patrons who contribute reviews, lists and other content in high quantity or with high quality
- Patron feedback

Each library is provided with a Google Analytics account for the library’s BiblioCore catalog in order to track statistics, search patterns, and other usage metrics.

BiblioCommons is a multi-tenant software service. That is, it is a cloud service with a single instance of the software that serves multiple libraries. As a result, their focus is on providing configuration options within the software, not customization of the software.

BiblioCommons can replicate the library’s header (logo, colors, and fonts) as well as full navigation. For consortia, BiblioCommons can use the consortium header and navigation or customize per library. For more customization, BiblioCMS can be used, which is a content management system that allows libraries to build and manage their website.

Patrons and staff can share and promote titles through Twitter, Facebook, and other social network sites through the use of the AddThis plugin (http://www.addthis.com/). Patrons can also review and rate items. These can be viewed for all users of BiblioCore system (catalog), not just the local users.

BiblioCore (catalog) pricing is based on the library’s legal service area; BiblioCommons pricing is sold based on flat formulas. In addition, there is a one-time implementation fee.

Current development includes:

- BiblioDigital, an integrated platform for reading, borrowing, and purchasing ebooks (currently in beta).
- BiblioEvents, which integrates events into the catalog and displays programs/events relevant to a patron’s search (expected in the fourth quarter of 2014).

Libraries using BiblioCommons west of the Mississippi River include:

- Austin Public Library: http://austin.bibliocommons.com/
- Daniel Boone Regional Library: http://dbrl.bibliocommons.com/
- Tulsa City County Library: http://tccl.bibliocommons.com/
- Pueblo City County Library: http://pueblolibrary.bibliocommons.com/
- Omaha Public Library: http://omaha.bibliocommons.com/

BLACKLIGHT

Blacklight is an open source Ruby on Rails Engine plugin that provides a discovery interface for Solr indexes. Blacklight provides a default user interface which is
customizable via the standard Rails (templating) mechanisms. Blacklight accommodates heterogeneous data, allowing different information displays for different types of objects.

Initial development was geared toward academic libraries, and deals with library data in MARC format. However, it is not limited to working with just MARC, but accommodates heterogeneous data and allows different information displays for different types of objects.

The primary functionality for Blacklight includes:

- Stable URLs for search and record pages allow patrons to bookmark, share, and save search queries for later access.
- RSS and Atom responses of search results.
- For certain types of solr documents, an OpenURL/Z39.88 COinS object is embedded in each document. This allows plugins like Zotero to extract data from the page.
- Support for OpenSearch, a collection of simple formats for the sharing of search results.
- Relevance-based searching with the ability to locally control the relevancy algorithms.
- Facets.
- Search queries targeted at specific sets of fields.
- Results sorting.
- Bookmarkable items.
- Permanent URLs for every item.
- Tools for exporting records to RefWorks or Endnote, sending records via email or SMS, or as a formatted citation.
- User tagging of items.

Blacklight does not:

- Have a central index.
- Require patron authentication; but if needed, can work with most providers.
- Provide statistics natively, but can work with web-based providers like Google Analytics.

There are technical requirements for implementing Blacklight:

- Ruby 1.9 or greater
- Rails 4
- Java 1.6 or greater
- Apache Solr
- Cascading Style Sheets
- HTML

As it is open source, the library can modify branding and functionality to fit its needs.
Relevancy can be modified by the library and is created by defining a multi-layered sort, e.g., sorting results first by the score field, then by the publication date, then by the title.

Although not part of the software, social media can be integrated using plugins such as AddThis. Tagging, reviewing, and rating items is not available in the software.

Current development focuses on:

- Statistics gathering
- Adding autocomplete functionality to the search form
- User interface enhancements
- Dropping support for older versions of Ruby and Rails

No libraries in Texas have been found to have implemented Blacklight; however, below are several outside the state:

- Columbia University: http://clio.columbia.edu/
- Indiana University - http://iucat.iu.edu/
- Johns Hopkins: https://catalyst.library.jhu.edu/
- Stanford University: http://searchworks.stanford.edu/
- University of Wisconsin-Madison: http://search.library.wisc.edu/

**EBSCO Discovery Services**

EBSCO Discovery Services, or EDS, entered the discovery tool landscape in 2010 as a response to Google search with the promise of better search results and user experience. EDS, according to EBSCO, differs from other discovery services because it offers both full-text searching and superior indexing of library resources. EDS works with major integrated library systems (ILS), and EBSCO is actively pursuing partnerships with others as their focus is to have a discovery tool “which works seamlessly in as many library environments as possible, regardless of which ILS or next generation library services platform a customer has chosen.” (Kelly 2013, p.36) By partnering with ILS vendors, EDS provides choices for libraries, because they can decide which interface to use, either the ILS or EDS platform, and also get more catalog functionality, such as view book availability and book checkouts. According to their website, EDS has partnered with these major ILS vendors: OCLC, SirsiDynix, and Innovative Interfaces (III).

EDS provides access to publisher-provided metadata from several subject indexes: Art Abstracts, ATLA Religion Database, Business Source, CINAHL, EconLit, Historical Abstracts, Inspec, PsycInfo, RILM Abstracts of Music Literature, and SocINDEX. Its content derives from several providers: Web of Science, Scopus, PubMed, JSTOR, ARTstor, LexisNexis, HeinOnline, AP Images, NewsBank and many others. Its full text search covers databases, e-books, e-journals, and e-packages ranging from magazines, journals, and trade publications, books, conference proceedings, CDs and DVDs to newspapers and newswires.
EDS works with link resolvers, which means EBSCO will automatically load your existing knowledge base data from other systems without the need for manual upkeep or redundant maintenance. Searches yield links to full-text which complement the library's link resolvers. Additionally, EBSCO does not require that EDS patrons work solely with EBSCO's knowledge base, but it synchronizes information automatically with other vendors’ knowledge bases.

In 2011, researchers at Bucknell University and Wesleyan University compared the search efficacy of several discovery tools: EDS, Summon, Google Scholar and conventional library databases. In terms of quantitative benchmarks, EDS outperformed Summon and the other search systems, and when evaluated by the librarians, the resources located by students using EDS were “judged as having a higher average quality than any of the other search systems tested.” (Asher et al., 2012, p.468) Also, the time to complete all four required searches was shorter when students used EDS rather than the other discovery tools.

The EDS relevancy ranking algorithm evaluates content by type, but also weights article lengths, which means that newspaper articles rank lower than journal articles, which are generally longer. Patrons can set up their own search parameters limiting searches both in time and scope. Another important finding in the Bucknell study was the patron’s over reliance on the discovery algorithm. Most of the students in this study did not search beyond the first page of results. (Asher et al; p.474)

Many libraries in Texas have implemented EDS; below are several examples:

- Texas Tech University
- Lamar University
- Texas A&M University
- Abilene Christian University
- The University of Texas at El Paso
- Brazoria County Library System
- Tarleton University
- Several core members of the Abilene Library Consortium

Two institutions from above, Brazoria County Library System and Tarleton State University, are highlighted by EBSCO as customer success stories in their website.

Since implementing EDS in 2013, Tarleton University has increased library usage of full-text downloads, has found that a single search helps both students and faculty find a broad range of quality peer-reviewed resources, and the ability to email and save records to patron’s personal folders facilitates broader collaboration.

For Tarleton University EDS patrons, Guided Style Find fields and helpful limiters are appealing and the familiarity of the interface combined with Tarleton’s branding allows for a consistent user experience throughout the discovery process. Patrons can access library resources from off campus via EZproxy, which requires patrons to log in before they access search results.
Brazoria County Library system implemented EDS in March 2014 as a single-search platform and they report an increase in circulation of both physical and electronic resources. Through EDS, they incorporated catalog records from Polaris, EBSCO databases, NoveList, NoveList Select, and LibraryAware. The relevancy ranking tool which elevates catalog records over other EDS content is particularly useful to Brazoria staff.

**EX LIBRIS PRIMO**

Similar to EDS, Primo was first launched in 2010 and offers several core functionalities:

- Fast response time for searches
- Relevance ranking
- Single point of discovery for all resources
- Collaboration capabilities that encourage patrons to add content such as reviews, ratings, and tags
- View content contributed by other patrons alongside a simple and intuitive user interface

The user interface in Primo is customizable. The goal is to maintain user familiarity throughout the discovery process.

Relevancy ranking is one of the key selling points for Primo. In order to overcome seeing false positive or irrelevant results high in the results list, Primo tries to balance all results with only relevant results data sets. They partnered with a company specializing in relevance ranking algorithms and tailored Primo’s algorithm to the library environment. On their website, Primo boasts a search response time for most sites below 500 milliseconds for an average search.

Primo is offered alongside its Central Index which is an aggregation of scholarly resources such as articles, e-books, reviews, and legal documents, which are harvested from primary and secondary publishers and aggregators, as well as several open source repositories. Several publishers have partnered with Ex Libris to enable the Primo Central Index to search their content or to extend the content that is already available for Primo Central Searches. Some of these content providers are Taylor & Francis, Project Muse, SAI Global, Palgrave Macmillan, Plunkett Research, CQ Press, Versita Publishing, and Bridgeman Education.

Another feature of Primo allows patrons to tag, review, and rank their search results. Both ratings and reviews of a resource appear in the full-item view; tags appear in the tags page for all items.

According to Ex Libris, there are two elements that differentiate Primo from other discovery tools:

1. Its open platform architecture. Primo includes an API for 50 services, which allows for customization and enables libraries to develop code extensions to share with other members.
2. Its deep search architecture. This enables deep searching, which means libraries can run Primo on top of external search engines. The results are ranked by relevance and displayed using faceted categorization.

Several libraries in Texas use Primo as their discovery tool; examples include:

- Midwestern State University
- The University of Texas at Dallas
- Texas Woman's University
- Texas A&M University

More recently, Harvard Library announced they will implement Primo in September 2014.

Midwestern State University went live with Primo in 2011. One of their primary considerations in choosing Primo concerned content neutrality, e.g., the discovery tool did not rank its own datasets higher than other library resources. They were already a Voyager customer and had experience with Ex Libris. They customized their interface and maintained familiarity with logos and banners for their patrons. (Fernandez, 2011)

Midwestern’s faculty and staff have responded positively to Primo. Midwestern’s Moffett Library reports their patrons utilize e-Shelf, which allows patrons to save searches and be notified via email or RSS when new relevant material is available. Patrons can also initiate ILL transactions and read articles across mobile devices. The library’s usage of previously overlooked library resources has dramatically increased since implementing Primo.

Harvard Library made the decision recently to adopt Primo as their discovery and delivery system. Their working group evaluated several discovery systems using these criteria: ease of use, coherent aggregation of local, licensed and open metadata, including non-textual and grey literature, support for interdisciplinary research, and flexibility both in the short and long terms as discovery system and accessibility continue to evolve. Additionally search precision, known-item retrieval, and integration with resource delivery and fulfillment systems were also functionalities the group used in evaluating discovery systems.

Primo features that can be attractive to libraries include:

- Seamless user experience, i.e., discovering an item, saving the result, making requests, viewing the account and renewing items
- Support for search and display of vernacular scripts and transliterations, i.e., Japanese, Chinese, Korean, Arabic, Hebrew and Cyrillic
- Browse searching for author, title, subject and LC call number, and phrase searching
- Exposing data to web crawlers for search engines like Google and allowing extensive configuration options
INNOVATIVE ENCORE

Encore is an Innovative Interfaces product and was first released in 2010 alongside Primo from Ex Libris and EDS from EBSCO. Unlike EDS and Primo, Encore does not have a pre-harvested index of content. Instead, article content is pulled for Encore search results in real-time using web services. (Rowe 2011, p.12) More recently, Encore has partnered with EBSCO to provide patrons with a broad collection of full-text articles, and an index which spans thousands of participating publishers and partner resources.

Encore also partners with Overdrive and 3M, allowing e-book integration and making the user experience seamless in terms of discovering both print and electronic items and showing real-time availability. Patrons can initiate e-book checkouts and holds for 3M materials from the Encore interface and view the status of these materials in their own browse and Encore account view.

Another partnership which enhances the user experience is through ChiliFresh, which encourages social interaction between the patrons and the library through a database of trusted ratings and reviews written by library patrons.

Some of the key features of Encore are single search results; integration of articles, books, e-books and digital collections; real-time ILS/LSP integration; and 3M Cloud Library, ChiliFresh, and Overdrive accessibility.

Encore can work with ILS products other than Millennium, which is developed and supported by Innovative. The interface can be customized by customers, in a manner similar to EDS and Primo.

There are no case studies of libraries in Texas that have successfully implemented Encore, nor is there published research since 2011 which can offer more insight on the implementation of this discovery tool.

OCLC WORLDCAT DISCOVERY

The newest entrant into the discovery services market made its debut in early 2014 - OCLC’s WorldCat Discovery. It is a suite of cloud-based applications that brings together the FirstSearch and WorldCat Local services. It enables people to discover more than 1.7 billion electronic, digital, and physical resources in libraries around the world.

The central index contains:

- 319 million cataloging records contributed by member libraries of OCLC
- 200 million citation records from ArticleFirst, MEDLINE, ERIC and other sources made available freely worldwide
- Over 1.1 billion article citations from licensed content providers and open access collections

Sources of this metadata include:
Discovery Services: A White Paper for TSLAC

- Member libraries of the OCLC cooperative
- Over 1300 institutional repositories from around the world
- Nearly 350 publishers, aggregators and societies

Currently, major formats included are:

- Over 250 million books
- Over 15 million e-books
- 12 million serial titles
- 16 million sound recordings
- 12 million visual materials
- 40 million digital items
- Over 4 million maps
- 7 million musical scores

Discovery provides access to both citation-based and full-text materials, as well as open access and public domain material, including HathiTrust, the Public Library of Science (PLoS), the Internet Archive, and Project Gutenberg. PubMed, PubMed Central, and SciTech Connect are also included in the central index, along with other .gov documents.

OCLC does not focus on licensing material provided to libraries, but rather partners with publishers, aggregators and societies, to receive and index their metadata into their central index.

If a database is not part of the central index, OCLC has two approaches to making it available:

1. When possible, OCLC negotiates with the vendor/publisher to load the data centrally at OCLC.
2. When that is not possible, they rely on Z39.50 to search the database remotely.

By August 2014, libraries will be able to include remote databases for which they have a subscription. When using WorldCat Local, these databases are searched on the content provider’s site, but results are displayed along with results from the central index. The assumption is that WorldCat Discovery will function similarly.

Libraries can choose not to display material from specific vendors or journals. They can also choose not to display full-text, but citations only. For example, when a patron conducts a search, the results will display brief information on the item/article. The patron must choose the “Full-text” link in order to view the article. The library can opt out of displaying this link.

The library can configure WorldCat Discovery to initially sort by just relevance, by relevance plus forcing the library’s holdings to the top (assuming the library has holdings in WorldCat), and by date, author, or title. A patron can re-sort results based on these same criteria; otherwise, they cannot change the relevancy.

Once the initial sort is completed, there are then several components to the WorldCat Discovery relevancy algorithm:
• Search terms in the author then title fields are weighted first, then title followed by the remaining fields in the record
• Term frequency
• Proximity of the terms to one another
• Currency
• How widely held
• Works in the language of the patron’s browser are elevated in ranking

WorldCat Discovery does not distinguish between material types for returned search results. However, within electronic resources, a library may decide to give a higher rating to a selected provider for an e-resource. This means for a single item from multiple providers, the link to the preferred provider will be sorted to the top of the list.

Patrons can view and export citations in the following formats: RefWorks, EndNote, EasyBib, HTML, RTF, and RIS. Patrons can export using RIS for exporting to Zotero and similar tools. The patron can generate citations in the following styles: APA, Chicago, Harvard, MLA, and Turabian.

Options for branding the interface include:
•Branded interface with logo and color options for banners, link text, and search/fulfillment buttons
•Configurable button text for links to link resolution, place hold/reserve, and request via interlibrary loan
•Custom links from WorldCat Discovery to related web sites and pages of your choice
•Branded version of the WorldCat search box for use on your Web site and other Web sites that are familiar to your patrons

Although WorldCat Local offers social media integration, WorldCat Discovery does not at this point. Its target date is listed as “to be determined.”

Configuration options include:
• Definition of institution, branches and/or group to determine the priority position in search results.
• Interoperability with a library’s local system in areas such as display of item availability, circulation and resource sharing, and OpenURL resolver.
• Definition of the workflow for placing institution requests, group requests and requests made to other WorldCat libraries.
• Configure available databases, database groups, and default search databases for different groups of constituents.
• Apply local branding such as colors, logos, etc.

OCLC follows the Open Discovery Initiative and is very interested, but is not part of the NISO working group.

Data mapping (or re-mapping) is not possible.
Subscribers will have access to the same type of usage statistics they currently have for FirstSearch activity. In addition, WorldCat Discovery Services will work with Google Analytics, a free tool, to deliver traffic and access statistics for your library. Libraries may purchase access to Adobe Reports & Analytics (formerly known as SiteCatalyst) for more customizable usage reports.

Future functionality is provided at http://oclc.org/en-US/worldcat-discovery/features.html. Examples include:

- Item location and availability on brief results option
- Additional relevance sorting options
- "Did You Mean?" enhancements
- Remote database search option
- E-book availability and checkout
- Ability to share on social media
- Permanent user lists
- API access for WorldCat Discovery
- Linked data exposure
- Google Preview integration
- Ability to customize facets

WorldCat Discovery does not dictate the type of authentication that a library uses. Authorization differs based on the action the patron is executing. Authentication for searching licensed databases and accessing full text is handled through IP via a proxy server. Authentication into the local ILS is managed by the library’s identity management service.

Many libraries are currently testing the WorldCat Discovery service. Below are two that are using it with their patrons:

- Texas A&M International University: http://library.tamiu.edu/
- Virginia Tech Libraries: http://virginiatech.on.worldcat.org/discovery

**PROQUEST SUMMON SERVICE**

For ProQuest Summon, discovery serves as a digital front door for the library, searching across the library’s resources and providing a starting place for research.

The Summon index contains over 1.8 billion records representing more than 90 different content types. All content in Summon is centrally indexed.

- All content in the Summon index is searchable at the same time with no reliance on federated search, XML APIs or other database platform technologies.
- The Summon service provides a single, unified result set.
- All content in the Summon service is treated equally. There is no bias in terms of relevance or content covered toward any one vendor platform or group of databases.
Content is mostly de-duplicated prior to indexing.

Local catalog records are pre-harvested into the index and MARC fields are mapped to the Summon index schema based on MODS. Libraries have control over this process and can submit custom-mapping criteria.

Institutional Repositories are harvested and ingested into the Summon Unified index.

Summon also supports Union catalogs and provides an “institution” facet that allows patrons to limit catalog results to a particular library or libraries’ holdings.

The Summon Service includes 516 open access and public domain databases, including institutional repositories and open access repositories. These include sites that provide metadata only, metadata with abstracts, metadata with full-text, full-text HTML, and full-text PDF. In addition, there are 1,117 databases and approximately 9,000 publishers available in the index. Libraries can display or hide materials from specific vendors or databases.

For libraries that have subscriptions, these can be separately tracked within the Summon Service and turned on, e.g., become discoverable, by the library’s patrons. Discovery of items that require ILL is also possible by selecting to see results beyond the library’s collection.

The Summon Service uses two scores, the Dynamic Rank and Static Rank, to define relevancy. Dynamic rank focuses on the query and is the more important of the two. It consists of:

- Proximity
- Term frequency
- Inverse frequency
- Field weighting
- Term stemming
- Stop word processing
- Synonyms
- Language processing
- Free-form identifiers
- Cut-and-paste excerpts

Static Rank focuses on the item itself and consists of:

- Content type
- Scholarly/peer-reviewed
- Publication date
- Citation counts
- Local collections
- Content size

This algorithm boosts local content so that they are more easily discoverable by patrons. Other resources cannot be privileged in this way.
Because the Summon Service does not index content by database or package, but rather matches and merges content at the item level, they do not provide facets based on database.

Libraries can map (or re-map) metadata for local content only. Metadata and full-text from publisher content is managed by ProQuest.

Patrons can create lists of search results including the ability to email, print, modify citation format, and export citations directly into Endnote, RefWorks, Zotero and other bibliographic management tools.

RSS and persistent URLs allow patrons to embed links in social media tools relatively easily. There is no native way for patrons to tag, rate, or review citations, although a library could add this functionality using the API.

The Summon service provides a Summon customizer tool for simple administrative tasks such as changing logos, choosing default languages, adding/deleting and arranging facets, as well as activating certain optional features (such as auto-complete and contextual facets). It also enables libraries to match their library look and feel. The Custom Translation Editor enables each library to customize all text on their interface, in any language.

The local administrator has the ability to customize numerous aspects of the Summon display interface, including number of results, number, type and order of facets, language, Database Recommendations, Best Bets, custom linking, record prioritization, institutional facet whitelisting, Union Catalog participant record prioritization institutional, and branding.

ProQuest is a member of the Open Discovery Initiative Working Group, as well as the KBART Working Group.

Summon Analytics tracks traditional metrics such as number of sessions and number of searches, but it provides in-depth behavioral analysis and user profiling reports to show libraries exactly how patrons interact with the Summon service. Search queries are also recorded allowing libraries to see the top search trends as well as track queries that may return few results. Summon provides easy integration with Google Analytics for libraries that prefer that or wish to supplement Summon Analytics with additional usage and user behavior metrics data.

Summon-supported authentication methods are as follows:

- OCLC's EZProxy
- Innovative Interface's WAM
- Any web-based authentication proxy that uses prepending URL rewriting to support IP-based authentication
- An institution wide VPN

The Summon Service uses campus IP authentication and integrates with various identity data stores.
Libraries within Texas that use Summon include:
- The University of Texas - Austin
- The University of Texas San Antonio
- The University of Texas - Arlington

**VuFind**

VuFind is a discovery layer and search engine. It is customizable and capable of presenting multiple data formats in a single user-friendly interface. VuFind is currently supporting thousands of libraries, museums, and archives world-wide.

VuFind is an open source PHP library search engine that allows patrons to search and browse catalogs as well as other databases. Created by Villanova University in 2010, it operates with a simple, Google-like interface and offers flexible keyword searching. The software is also modular and highly configurable, allowing implementers to choose system components to best fit their needs.

The latest version, 2.2, includes new themes using the Bootstrap framework which adds responsive design features. Primary functionality for VuFind includes:

- Search with faceted results
- Live record status and location (VuFind queries the ILS)
- “More Like This” suggestions
- Save resources to lists
- Text or email citations
- Tagging and commenting
- APA and MLA citations
- Author biographies
- Persistent URLs
- Zotero-compatible
- Internationalization
- OpenSearch, Open Archives Initiative (OAI), Solr

VuFind does not:

- Have a central index
- Require user authentication, but if needed, can work with LDAP, MultiAuth, SIP, Shibboleth, CAS

There is a statistics module, but it is difficult to determine what type of statistics it provides. Google Analytics can be used with VuFind.

There are technical requirements for implementing VuFind:

- Apache HTTP Server 2.2 or greater
- PHP version 5.3.3 or greater
- MySQL 4.1 or greater
Java J2SE JDK 1.6 or greater
Windows or Linux operating systems
Cascading Style Sheets
HTML

As it is open source, the library can modify branding and functionality to fit its needs.

Relevancy can be modified by the library and is created primarily by changing the weighting and fuzziness of specific types of searches.

Although not part of the software, social media can be integrated using plug-ins such as AddThis. Tagging and commenting are available in the software.

Current development is focusing on:
- Updating translation files
- Create a plug-in architecture
- Implement standard model for consortia
- Update breadcrumb navigation
- Implement FRBR
- Implement recommendation system
APPENDIX B

WHAT TO LOOK FOR IN A DISCOVERY SERVICE CHECKLIST

☐ Usage Statistics Functionality (minimum):
  ☐ Total number of searches
  ☐ Result clicks
  ☐ Total number of click-throughs
  ☐ Total number of searches per month
  ☐ Total number of unique visitors per month
  ☐ Total number of click-throughs per month
  ☐ Top 500 search queries for the last period
  ☐ Top 100 referring URLs to the discovery service for the last period

☐ Relevancy Ranking Practices:
  ☐ Ascertain how relevancy is determined; can it be modified by the library?
  ☐ Is the ranking of search results objective?
  ☐ Can specific providers or types of documents be privileged?
  ☐ Does the relevancy ranking algorithm provide results helpful for your patrons?

☐ Central index:
  ☐ Which databases are available in the central index?
  ☐ What is the quality of the metadata?
  ☐ Does it include the types of materials you need, e.g., full-text, citations, journal backfiles?
  ☐ Is the full-text searchable?

☐ Discovery layer:
  ☐ Does it include advanced searching options, facets, and limiters? Are they easy to understand and use?
  ☐ Does it include end-user features helpful for your patrons, e.g., lists, tagging, citation export, and social media integration?
  ☐ Can you customize the look and feel or branding of the website? Are widgets and APIs available?
  ☐ Can results be enhanced with cover art, recommendation engines, or other external information?
  ☐ How usable is the site for patrons?

(From Open Discovery Initiative and Hoeppner 2012)
APPENDIX C

EVALUATING YOUR DISCOVERY SERVICE CHECKLIST

☐ Evaluate decisions made during the implementation process.
☐ Utilize patron feedback to inform and ameliorate future workflows.
☐ Analyze the goals of implementing a discovery service. Were they achieved?
☐ Vendor Company:
  ☐ Understand the contractual agreements between you and the vendor.
  ☐ Understand the type of support you will receive from the vendor as part of your agreement; evaluate that experience. Focus on:
    ☐ Company stability
    ☐ Quality of their staff
    ☐ Your experience during support interactions (outside the implementation process)
    ☐ Quality of help they make available
    ☐ The process through which they handle conflict
    ☐ How they handle system updates and technical support
☐ Resource Coverage and Indexing:
  ☐ Measure resource usage against what the vendor advertises
  ☐ Work alongside your vendor to understand what items are discoverable
  ☐ Perform searches that cover full text, subject headings, and abstracts
  ☐ Review search default settings for your institution
☐ Usage Statistics:
  ☐ Through this data, which portion of library resources are your patrons finding through the discovery service?
  ☐ What types of information are they accessing through other resources?
☐ Relevance Rankings:
  ☐ Understand which resources are rising to the top of their search results for given searches
  ☐ Run searches to discover if the vendor’s own data will appear at the top of the search results first or a combination of sources
☐ Testing:
  ☐ Identify the resources and tools through which you will gather testing data
  ☐ Build your network of resources, e.g., other colleagues who are using the same system, and solicit their advice when building test scenarios
  ☐ Maintain a working relationship with your subscription vendor and utilize their help in connecting you to other sources in your region
  ☐ Build your own scenarios based on patron feedback
  ☐ Build survey forms that display after patron search sessions
Follow-up with patrons via telephone or face-to-face interviews to capture the user experience.

Rely on focus group feedback to determine enhancements to the services.

**Trends:**
- Be aware of new developments or enhancements to your system.
- Collect information from the vendor, e.g., technical information, case studies from other institutions with the same patron demographic.
- Attend seminars, conferences or other on-going focus group meetings.
- Participate in focus groups that seek to improve your system’s operation.
APPENDIX D

BIBLIOGRAPHY (2011-PRESENT)


