The Future of Library Resource Discovery

A white paper commissioned by the NISO Discovery to Delivery (D2D) Topic Committee

By Marshall Breeding

February 2015

About NISO White Papers

NISO White Papers are contributed or solicited papers that address an issue that has implications for standards development. White Papers can be viewed as a pre-standardization activity. A NISO White Paper might define and explore some of the questions that come into play before formal standardization work is started. Or, a NISO White Paper might identify areas that are opportunities for standards development and suggest possible approaches. All White Papers are posted on the NISO website (www.niso.org).

Published by:
NISO
3600 Clipper Mill Road
Suite 302
Baltimore, MD 21211

ISBN: 978-1-937522-41-4

Contents

Introduction

Libraries have a strategic interest in the tools and technologies that facilitate the discovery of and access to the resources for the communities that they serve. These tools have seen steady advancement over recent decades, making great strides in the scope and depth of materials addressed and in providing library users ever more convenient ways to access these materials. The progress seen in the successive generations of technology beginning from online catalogs, to metasearch tools, to the current generation of index-based discovery services represents an incredible improvement. Yet many gaps remain relative to the potential of more universal access to the universe of content of interest to libraries and their users and obstacles remain that impede progress that NISO or other organizations can address.

This paper provides an overview of the current resource discovery environment and discusses some of the possibilities regarding how these technologies, methodologies, and products might be able to adapt to changes in the evolving information landscape in scholarly communications and to take advantage of new technologies, metadata models, or linking environments to better accomplish the needs of libraries to provide access to resources.

The paper will include recommendations that can be addressed in the short term through a possible extension of the NISO Open Discovery Initiative as well as longer-term efforts that investigate how evolving technologies such as open linked data can be operationalized. Areas of possible activity include: a second phase of the Open Discovery Initiative that tackles topics considered out of scope in its initial phase; facilitating more cooperative development of the application program interfaces (APIs) that comprise the ecosystem among discovery services, resource management systems, learning management systems, and other components in a campus information environment; promoting additional research in the extension of index-based discovery services to take advantage of the emerging realm of open linked data.

The paper will also attempt to look beyond the current model to explore other alternatives, especially related to the realm of linked data. To what extent can we expect that the universe of content of interest will be available as open linked data? What technologies would facilitate improved user experience based on linked data? What new capabilities would be possible?

This paper does not aim to make detailed comparisons of products, but to observe that this product genre continues to experience significant momentum in its development trajectory. Areas of interest or concern for this white paper include exploring or identifying the factors that cause significant barriers to progress, if there are methods with the potential to improve discovery that have not been adopted by the current field of discovery services and need additional exploration or stimulation, or if there are opportunities to lower thresholds of entry to enable additional organizations to contribute to the discovery services arena.

1. General Background

This investigation addresses the broad topic of the methods and technologies available to libraries to make their resources discoverable and accessible by the communities that they serve. It will examine both the tools created or acquired by the library for its patrons to use directly as well as those that may operate outside of the environment directly controlled by the library.

The current discovery environment in the academic library arena is dominated by a set of products within the genre of index-based discovery services, often marketed as "web-scale discovery services," which rely on a large central index populated by metadata, full text, or other representations of the content items in a library's collection. These indexes are a component of a multi-tenant platform comprised of search and retrieval technology components, and an end-user interface. The platform may also expose APIs that allow programmatic access to the search and retrieval functionality of the platform that bypasses the provided interface. This group of discovery services does not exist in isolation, but as part of the ecosystem of scholarly and popular publishing, abstracting and indexing (A&I) services, and in the information infrastructure of the libraries that adopt them.

a. Discovery Components and Categories

The arena of library-provided resource discovery products includes several different categories. Each of these categories addresses a specific scope of functionality and underlying components.

Discovery interface

Discovery interfaces, originally marketed as "next-generation catalogs," emerged to provide a more modern replacement to online public access catalog (OPAC) modules of integrated library systems (ILS). They provide an improved end-user interface used by researchers to submit queries, receive results, and make content selections. A discovery interface includes features such as relevancy-based search results, faceted navigation, and other features consistent with web-based resources and these multiple areas of functionality:

- **End-user interface,** usually delivered via a web browser, to perform tasks such as presentation of a search box for end-user queries, an alternative query page that presents advanced query options able to target terms according to structured fields, and presentation of search results listed either in a brief form or in full-record displays.
- **Interoperability with a link resolver** to present links to full text from citation records in search results.
- **Local search and retrieval,** usually through an integrated indexing, search, and retrieval component to collections of interest. Many local search and retrieval indexes use Apache SOLR™ (lucene.apache.org/solr/) or ElasticSearch® (www.elasticsearch.org/) as the local search tool.
- **Ability to interactively communicate with the library's ILS implementation** for tasks such as determining the current availability status of items in the library's physical collection, to transmitting requests for holds or recalls, and interacting with the patron records to present current account status, lists of items charged, fines or fees due, and to view or update personal details.

- **Access to remote index platforms via API** in addition to, or instead of, targeting search queries and receiving results from a local index. A discovery index may also operate directly with an external platform that indexes content of interest. This interoperability is made possible through a mutually defined set of APIs that manage the requests, responses, record transfer, and document presentation needed to support a search session.

Both open source and commercial products have been created in this category of discovery interfaces. Commercial examples of discovery interfaces include:

- **Ex Libris® Primo®** was originally developed in 2006 as a new-generation interface to provide relevancy based search for materials managed by a library's integrated library system, local content repositories, and other collections of interest that might be available for local indexing via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) or batch record loading. Ex Libris later developed Primo Central as a managed central index of scholarly content. Primo uses Apache Lucene or SOLR as its technology for managing local indexes.
- **SirsiDynix® Enterprise®** provides a platform for relevancy-based retrieval and faceted navigation for the content managed in a library's integrated library system. **SirsiDynix Portfolio™** extends the platform with a digital asset management component that enables a library to describe local digital materials that can then be searched through Enterprise.
- **BiblioCommons** provides **BiblioCore**, which includes hosted discovery service relevancy-based retrieval, faceted navigation, and a variety of social and community-oriented features. BiblioCommons maintains a discovery index that includes an aggregation of its customers' records from their respective ILS implementations.
- **ProQuest® AquaBrowser® Library** provides an end-user index with faceted navigation and a cloud of search terms extracted from search results that can be used to execute new searches. AquaBrowser maintains a local index based on proprietary technology populated by records extracted and synchronized from the library's ILS implementation.
- **Innovative Interfaces Encore**, originally introduced in 2006, supplements or replaces the online catalog of Millennium or Sierra with a new interface that features a single search box, faceted navigation, and results ordered by relevancy. Encore was originally designed to operate with automation systems other than those from Innovative, but saw very few implementations. Encore is currently only used by libraries using Millennium or Sierra. Encore does not include its own article-level discovery index; Innovative works with EBSCO Information Services to integrate EBSCO Discovery Services™ for libraries that subscribe to both products. Encore also includes integration with e-book lending platforms.

Open Source examples of discovery interfaces include:

- **Blacklight**, originally developed by the University of Virginia, is based on a Ruby on Rails programming framework and Apache SOLR indexing, search, and retrieval technology. Blacklight provides a flexible toolkit for a wide variety of record types and is the predominant search interface to the Hydra Project digital asset management system.
- **VuFind**, originally developed at Villanova University, is based on a PHP programming codebase and Apache SOLR indexing search and retrieval technology. VuFind has been implemented in thousands of libraries. Many of the projects work with forks of the original codebase and the development efforts can be characterized as independent and fragmented.

- **eXtensible Catalog**, "a research project launched in April 2006 by the River Campus Libraries of the University of Rochester, with funding from the Andrew W. Mellon foundation, has created a number of tools that complement the development of discovery products and services. The main outcomes of the project include a set of connectivity tools, including toolkits for the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and for NISO Circulation Interchange Protocol as well as the XC Metadata Services Toolkit. This toolkit offers utilities for the transformation and clean-up of metadata as it is extracted from repositories, such as library management systems, and loaded into discovery services. The eXtensible Catalog project has also created the XC Drupal Toolkit that provides a discovery interface with customizable faceted navigation based on content from repositories and the library website. Though the toolkits created by the eXtensible Catalog have been used by many projects, no libraries have yet placed the full set into use as their primary discovery interface." [Source: Breeding, Marshall (2013).] Update: The Kyushu University Library in Japan has implemented a discovery interface for its local catalog holdings based on the XC Drupal Toolkit of the eXtensible Catalog. (catalog.lib.kyushu-u.ac.jp/en)
- **Franklin** is a local discovery interface developed by the University of Pennsylvania Libraries, which is not based on Blacklight or VuFind.

Index-based discovery services

Index-based discovery services include a discovery interface with the characteristics described above, but which also provide a central index populated by resources that represent the general body of content of interest to libraries. These indexes are massive and aim to cover the body of academic library-oriented content (or a specific subset of content, resources, and services available on the Web).

The central index of these discovery services is potentially generated from a variety of categories of content, including:

- Metadata and full text from commercial publishers
- Content from A&I resources
- Metadata and full text from open access repositories
- Metadata or full text from relevant institutional repositories
- Bibliographic and holdings information from a library's resource management system

Material from publishers is generally included by virtue of an arrangement where content is moved in bulk through some arranged technical transfer mechanism. The full text of content items is often provided to generate index entries. Unless the discovery service provider either owns the rights to the full text or has licensed access, users will be linked to articles on the publisher's servers rather than be served the copy used for indexing.

The current arena of index-based discovery services is entirely dominated by commercial products. The major products include:

- Primo and Primo Central from Ex Libris Group
- EBSCO Discovery Service from EBSCO Information Services
- Summon® from ProQuest

- WorldCat® Discovery Service from OCLC®

To date, no open source index-based discovery services have been created based on an open access or community created central index. Section 8.e discusses potential opportunities in this area in more detail.

These index-based discovery services continue to evolve in a highly competitive commercial arena. Each of the products have seen a continual advancement through their product cycles to expand the content represented in their indexes, to add new features to their end-user interfaces, and to improve the performance of their relevancy or other search and retrieval capabilities.

The volume of materials in the body of content considered within the scope of these discovery services is vast. Counting the number of records represented in the index is not necessarily a valid way to characterize the completeness of the index, but helps to illustrate the magnitude of content that must be addressed. Not all of the providers of the commercial index-based discovery services mention publicly the size of their central index. ProQuest, however, stated on its site in January 2015 (www.proquest.com/products-services/The-Summon-Service.html): "Summon: The Summon service is the only discovery service based on a unified index of content. More than 90 content types, 9,000 publishers, 100,000 journals and periodicals, and 1 billion records are represented in the index. New content sources are added every week and content updated daily."

The index design of ProQuest Summon is based on a single record representing each unique resource, which merges records from different sources or providers for the same resource. This single-record strategy amalgamates citation, full-text, and A&I resources. Other products, such as EBSCO Discovery Service, maintain separate records from each source that represents any given resource. This difference in record strategy has many implications in the functionality of the index, but it also means that those that do not merge records may have substantially more than the 1 billion reported by ProQuest for Summon.

Local index content

The creators of discovery services populate their indexes with content from a variety of sources in addition to article-level scholarly content from proprietary and open access sources. The ability to incorporate local resources can have very high strategic value to the institutions implementing the service. The content for these local resources is harvested from a variety of sources including the institution's integrated library system, digital collection management platforms, or institutional repositories. The representation of these resources may need to be segregated in the indexes so that they are seen primarily by their own institution and not others that may be customers of a multitenant discovery service. Each of the index-based discovery services follows its own strategy in how it handles indexing and retrieval of local materials versus those represented in its global central index.

Through local indexing or other mechanisms, the discovery service can be used to provide access a variety of collections types and content that a library may manage on platforms other than its primary integrated library system. Some of these categories of content include:

- **Archival material** – Many libraries operate a separate system such as ArchivesSpace, Adlib Archival Management System or Calm from Axiell, or other commercial or open source products to manage their special collections or archives. These materials have distinct needs including specialized metadata formats.

- **Digital collections** – Libraries managing digital collections, such as digital or digitized photographs, manuscripts, video, or sound recordings usually operate a separate digital asset management system (e.g., OCLC CONTENTdm®, Ex Libris DigiTool® or Rosetta, Greenstone, etc). These collections may be of interest within any broad discovery tool that the library implements. Challenges apply to the potentially large number of items in these collections and how the system will perform the relevancy rankings so as not to overwhelm other materials.
- **Institutional repositories and electronic theses and dissertations** – These systems manage unique and important content for an academic institution on a separate platform. This content may or may not be also represented in the library's integrated library system. Discovery environments regularly harvest metadata or full text of these materials from the repositories via OAI-PMH.
- **Museum or exhibition materials** – These may also be managed in separate platforms such as the open source CollectionSpace (www.collectionspace.org/) or through the many different commercial museum management systems.

A variety of discovery configurations can be implemented for these locally managed collections. In some cases, the emphasis is on providing access only to the local community. Owner or copyright issues may apply that require local limitation of discovery. Each of the major discovery services offers some capabilities for local materials to be indexed in a separate way with appropriate access restrictions. In other cases, the collections are of global interest and can be included in the global and public version of the discovery service index. The partitioning of local from global content adds a layer of complexity to these services for metadata harvesting and resource access mechanisms.

Non-library discovery service

Researchers do not always make use of services provided by libraries. Services such as Google Scholar or Microsoft Academic Search can be seen as an alternative to the index-based discovery services produced by library-oriented organizations. Among the non-library scholarly discovery tools, Google Scholar dominates. Many researchers will also rely on the repositories, indexes, or other resources that prevail within their specific discipline. In disciplines with a well-established service, such as MEDLINE® for biological sciences or arXiv for physics, researchers will have less of a need for a library-provided discovery service to search that literature.

Google Scholar provides an index for scholarly materials that is widely used by students of all ages and researchers. The scale and sophistication of technology involved in Google Scholar far exceeds that of the library-oriented service providers.

Google ranks as one of the world's largest companies and does not provide detailed information regarding how it constructs or provides its services. The company's business model depends primarily on advertising revenue, but it does not display ads with Google Scholar results. Library-oriented discovery services depend on the subscription fees paid by libraries and must provide sufficient accountability, features, use statistics, and access to content to continue to sell their products. Google, in contrast, does not require payment for the Google Scholar service nor does it volunteer substantial information regarding its operational details.

Google also does not reveal the scale and completeness of the material available through Google Scholar. A variety of studies have been performed to better understand these details. A very rough

estimate suggests that Google Scholar indexes around 160 million articles. [Orduña-Malea, Enrique, et al. (2014).]

Nature published an interview in November 2014 with Anurag Acharya, one of the co-developers of Google Scholar, that provides some insight to the service. Acharya mentions that the Google Team currently includes nine individuals, but does not give details such as how many resources are represented in its index. In the interview, Acharya acknowledged that Google Scholar does not provide revenue, but is not expensive to operate and is perceived to provide significant benefit. When asked about the possibility of an API for access to Google Scholar results, Acharya's response provides some insight into the general model of the service:

> *Many people would like to have an API (Application Programming Interface) in Google Scholar, so that they could write programs that automatically make searches or retrieve profile information, and build services on top of the tool. Is that possible?*
>
> I can't do that. Our indexing arrangements with publishers preclude it. We are allowed to scan all the articles, but not to distribute this information to others in bulk. It is important to be able to work with publishers so we can continue to build a comprehensive search service that is free to everybody. That is our primary function, and everything else is in addition to this.
>
> [Van Noorden, Richard. (2014).]

Google Scholar operates though substantially different mechanisms than the library-provided discovery tools. The index for Google Scholar is populated primarily through automated processes. Rather than depend on bulk transfers of content from publishers, it makes use of the same kind of harvesting robots for scholarly resources as it employs for indexing the general web. Google operates a massive network of automated robot applications that scour the web for content. These bots are able to detect scholarly resources available on the open web and include them in the Google Scholar index. The indexing bots associated with Google Scholar systematically collect metadata and full text for all material that it considers to be scholarly. It is generally understood that Google Scholar has its own set of harvesting bots that are different than those for the general Google search, but there are likely synergies.

Google Scholar also has arrangements with major publishers to index proprietary content. These arrangements enable the Google Scholar harvesting bots to access documents within their secured servers that would otherwise be available only to authorized subscribers. This live harvesting facilitates a more current representation of recently published articles compared to the bulk transfers employed by the library-oriented index-based discovery services.

Google Scholar's indexing policies for dealing with publisher content (http://scholar.google.com/intl/en/scholar/publishers.html#policies) includes:

- **Publishers have control over access to their articles**
  We work with publishers to preserve their control over access to their content and only cache articles and papers that don't have access restrictions. Publishers can help us by identifying the regions of their sites that have access restrictions.
- **Google users must see at least the complete abstract or the first full page**
  This is a necessary component of our indexing program. For papers with access restrictions, all

users clicking on search results must see at least the full author-written abstract or the first full page of the article without requiring to login or click on additional links.

Google Scholar functions only as a search engine for scholarly articles. It does not provide access to any full-text content that would not otherwise be available to the searcher. Articles covered by library subscriptions can be linked directly, with others available via request through the library's document delivery or interlibrary loan service or through payments to the publisher.

Google Scholar provides mechanisms by which libraries can facilitate access to materials by those within their communities. Libraries are able to provide Google with the IP addresses of the campus and the base URL of its link resolver to provide easier access to subscription-based electronic resources accessed through Google Scholar. (See: https://scholar.google.com/intl/en-US/scholar/libraries.html)

Google Scholar certainly plays a very large role in the discovery arena. Roger Schonfeld suggests that it delivers researchers to resources more frequently than the library-provided discovery services [Schonfeld, Roger D. (2014)]. Several points may apply relative to Google Scholar when considering directions and opportunities for resource discovery:

- Libraries are not likely to have a large role in shaping the future of the service or gaining additional transparence regarding its details of operation or content covered.
- It is likely, but not guaranteed, that Google Scholar will continue to be available. Its future seems at least as promising as any of the library-oriented discovery services.
- Libraries can help expand the content indexed by Google Scholar and other search tools based on web harvesting. Resources and repositories managed by libraries can improve their exposure through techniques such as unique durable links, embedded metadata in page headers, resource content encoded with semantic structure such as through schema.org, listing of unique resources through the sitemap.org protocol, and other standard search optimization techniques.
- Zepheira, for example, provides services related to the improved performance of library resources through the incorporation of linked data.
- Techniques implemented that target improved performance in Google Scholar and other web search engines may also provide benefits to discovery services provided by libraries as the ecosystem evolves to include new services based on linked data.

Article-level discovery services not based on central indexes

While there has been a major shift toward reliance on central indexes in support of discovery and away from technologies such as federated search, the change is not universal. Some institutions and projects have made deliberate choices to not adopt the index-based discovery model. It is not clear which institutions have deliberately avoided investment in this model of discovery and which may just be later adopters.

Stanford University, for example, has opted not to implement one of the commercial index-based discovery services. Even as one of the top research libraries in the world, it has not seen a great deal of interest from its patrons in having them acquire one of the commercial products. Stanford depends on both its locally developed SearchWorks interface, access to individual databases and collections, and use of federated search tools from Deep Web Technologies for access to article-level resources.

One of the areas of interest for Stanford lies in increased opportunities for allowing users to navigate to results through browsing. Tom Cramer, Chief Technology Strategist and Associate Director for Digital Library Systems and Services provided the following comments in an e-mail exchange:

> We're looking at author/title and (to a lesser degree) subject browse based on authorities, as well as browsing based on LOD (linked open data). What we're particularly interested in is different user interactions than the traditional (and I would say, poor) ILS-ish browse of alphabetized lists. For example,
>
> - We've looked at auto-suggest and type-ahead (with values pulled from our records / authority files) as alternatives to the list interaction, for dealing with misspellings and/or transliterated names with alternative spellings. We plan to implement this.
> - The spoke and bubble visualizations for RDF / linked data that seemed so trendy a few years back (e.g., in AquaBrowser's sidebar) don't seem particularly effective or user friendly, as showy as they might be.
> - University of Wisconsin had an interesting version of subject browse that was FASTish, allowing patrons to effectively add components to a subject browse and compose their own multipart subject browse categories. It was beta at the time and they've since taken it down, but this is a good potential source for inspiration.
> - We might also look at something like DataTables jQuery plugin for browsing our authority lists. This presents very long lists of information but instantaneously filters the list down to match input in a search box—kind of like a hybrid between type ahead and browse queries with alphabetized lists. (See how we work with Course Reserves in SearchWorks for an example of this interaction. Try typing in "history", e.g.)
> - At some point, we'll probably add geographic browse / spatial search in, but this is not a top priority based on expressed user needs.
>
> So I guess what we're looking for is a site that has absolutely nailed "browse" in a way that leverages the richness of library data, but in a better set of user interactions than a traditional OPAC to meet the various needs.

There are other examples of academic libraries that have not implemented index-based discovery services. Utrecht University Library, for example, has opted to forgo library-provided discovery altogether and rely instead on general Internet-based tools.

Public library discovery services

Public libraries face much different issues in the realm of discovery than those that serve colleges and universities. While academic libraries devote the dominant portion of their collection resources to scholarly e-journals, public libraries continue to be engaged primarily with books—with e-books representing ever higher levels of interest. The discovery environment for a public library needs the ability to search local print collections, licensed e-book collections, modest collections of scholarly and popular electronic resources, as well as any local repositories of content.

In addition to the discovery services oriented to academic and research libraries, a variety of related products and services appeal to different types of libraries. Products oriented to public libraries include, for example, BiblioCommons and AquaBrowser. The online catalog modules of many of the integrated library systems have developed to a point where there is considerable overlap and convergence with discovery interfaces. These online catalog products, though generally tied to the vendor's own ILS products, increasingly offer local indexing capacity, faceted navigation, and integration options with

index-based discovery services. Examples include the ProPAC for Polaris, Encore from Innovative Interfaces and the LS2 PAC from The Library Corporation.

One of the top issues in the public library arena involves the ability to fully integrate e-book discovery and lending into the online catalog or other search interface provided. The initial phase of library e-book lending relied on directing patrons to the platform of an external provider if they wanted to access e-books. Today, there is an expectation that e-book lending should be available within the library's own interfaces. To meet this expectation, the online catalog needs to be able to include the content of the e-books currently available from external services and to be able to perform transactions through APIs or other mechanisms that allow patrons using the library interface to select, check-out, and download e-books into their devices.

Much progress has been accomplished toward a more integrated e-book experience for library patrons, especially as articulated by the ReadersFirst initiative (readersfirst.org/). Additional possibilities for development in this area would include an expanded and more standard set of APIs defined in the ecosystem of e-book lending platforms and discovery interfaces that would expand the functionality currently available programmatically and that would simplify the interconnections for implementations that involve multiple e-lending targets. Detailed analysis of the e-book integration into discovery interfaces is not considered directly within the scope of this study, but is an important area that warrants further investigation of opportunities for standardization.

While the general trajectory of discovery for public libraries differs from that for academic libraries, there are also considerable areas of overlap where both communities can benefit from development of recommended practices, development of open protocols, or other cooperative efforts.

Comprehensive library portals that include discovery

In the academic arena, discovery services have been developed as a component that would reside among other parts of a library's overall web presence. In other sectors, especially public libraries, a genre of interface products are available that not only include discovery capabilities, but also provide a content management system and other functionality that provide a complete replacement for a library's website.

These products provide a seamless presentation that unites discovery-oriented tasks with the many other activities that are supported through a library website. They have the potential for making all of the descriptive information about the library's services and programs, finding tools, and other content in a library's website more accessible and discoverable to library users.

Some of the commercial products in this genre include:

- Iguana from Infor Library Solutions
- Arena from Axiell
- BiblioCMS from BiblioCommons
- Enterprise from SirsiDynix (optional capability)

These products may have implications for how discovery services can be incorporated into a more integrated presentation layer provided to library users. These products are not themselves considered

within the scope of this report, but illustrate that the presentation of library resources and services, of which discovery is a subset, increasingly needs to be expressed through interoperable technologies that can be integrated into a unified presentation layer or as widgets that can be incorporated into third-party interfaces.

b. Selected Studies and Reports

The topic of resource discovery has been extensively covered in the library literature. Following increased levels of investment and adoption in discovery services, there are now opportunities to measure the impact of the specific products relative to use prior to implementation, to compare results among products, and to assess how they compare with non-library discovery scenarios such as Google Scholar. A few recent articles and white papers illustrate this conversation playing out in the literature.

Texas State Library and Archives Commission, with support from the IMLS developed *Discovery Services: A white paper for the Texas State Library and Archives Commission*. The work begins by providing definitions and a brief overview of the current state of discovery services. The paper highlights advantages of discovery services in enhancing the leverage of a library's collection of electronic resources in a more effective way than general Internet search engines or federated search tools. Limitations mentioned include lack of completeness, imperfect relevance ranking, and possible concerns regarding speed of cloud-based services. The white paper provides a checklist of what libraries should expect in a discovery service. The paper considers discovery interfaces and index-based discovery services. [Kabashi, Arta, et al. (2014).]

A variety of studies aim to demonstrate the impact that the use of a discovery service might have on the use of collections. An initial phase of a planned larger study, presented at the 2014 UKSG Conference, statistically measured the impact of the four major discovery services on collection, demonstrating general increased use, but with considerable variation. [Levine-Clark, Michael, et al. (2014)]

Other studies have focused on the impact of a given discovery service in a library. Calvert studied the impact of EBSCO Discovery Services on the collections at Northwestern Carolina University, noting "demonstrable increase in the use of abstracts and A&I databases…." [Calvert, Kristin. (2015).]. Lundrigan, Manuel, and Yan explored the level of satisfaction that users have with the Summon service implemented at Ryerson University. [Lundrigan, Courtney, et al. (2015)]. Asher, Duke, and Wilson presented observations regarding the results of three different discovery scenarios in multiple institutions on students' abilities to locate resources, noting a superior performance of EBSCO Discovery Service. [Asher, Andrew D., et al. (2013)]

Aaron Tay has done some interesting work analyzing how Google Scholar functions relative to Summon and other library-oriented discovery services. His June 2014 blog posting *8 Surprising things I learnt about Google Scholar* [Tay, Aaron. (2014).] and April 2013 posting *How are discovery systems similar to Google? How are they different?* [Tay, Aaron (2013)] provide an interesting summary of his observations.

The January 2014 issue of *Library Technology Reports* "provides an introduction in to the genre of library resource discovery products and provides product descriptions and perspectives on the major products." [Breeding, Marshall. (2014).]

Roger C. Schonfeld of Ithaka S+R authored a study that has received considerable attention. *Does Discovery Still Happen in the Library? Roles and Strategies for a Shifting Reality* reflects research that indicates the extent to which researchers discover materials in other ways, especially through Google Scholar, than through the discovery services implemented by libraries. Schonfeld argues for a "more integrated vision for discovery" that goes beyond the library-provided discovery service to include a more global approach that leverages tools provided by Google and cooperative efforts to make resources more discoverable. [Schonfeld, Roger C. (2014).] Schonfeld's study provides important context that an index-based discovery service chosen by the library will not serve all users at all times. These discovery services complement other tools that the library will provide including aggregated databases with their own search capabilities, either in general or for specialized disciplines, the online catalog for print resources, finding aids for archives and special collections, as well as search engines on the open Web, such as Google and Google Scholar.

Simone Kortekaas of Utrecht University Library describes considerations related to the need for a discovery service in *Thinking the Unthinkable: A Library without a Catalogue — Reconsidering the Future of Discovery Tools for Utrecht University Library*. Coming from a library that offered its users an online catalog for print resources and a separate interface for electronic materials, whether or not to implement one of the commercial discovery services was considered, but ultimately rejected. Utrecht University opted for complete reliance on Google Scholar, WorldCat, and other Internet resources for discovery since their research indicated that is what their users expected. [Kortekaas, Simone. (2012).]

A September 2014 white paper published by OCLC, *Success Strategies for Electronic Content Discovery and Access*, does not focus on discovery services, but on how resources can be made more discoverable through more accurate and complete metadata, and e-resource holdings knowledge bases that are more accurate and synchronized with library subscriptions. [Kemperman, Suzanne Saskia. (2014).]

c. Applicable Standards and Recommended Practices

NISO Open Discovery Initiative

The NISO Open Discovery Initiative Working Group (www.niso.org/workrooms/odi/) developed a recommended practices document for *Promoting Transparency in Discovery* that has been approved by the Discovery to Delivery Topic Committee and published as NISO RP-19-2014. It includes guidelines to content providers on disclosure of level of participation, the minimum set of metadata elements provided for indexing, linking practices, and technical formats. Recommendations for discovery service providers address content listings, linking practices, file formats and methods of transfer to be supported, and usage statistics. The document also provides background information on the evolution of discovery and delivery technology and a standard set of terminology and definitions for this technology area.

NISO has established a standing committee to maintain and support the recommended practice of the Open Discovery Initiative and address related issues.

NFAIS Recommended Practices: Discovery Services

The National Federation of Advanced Information Services' *Recommended Practices: Discovery Services* was published in August 2013 [NFAIS (2013)]. Addressing the interests of the providers of abstracting and indexing services, this set of recommended practices highlights specific issues and concerns that

apply to their potential cooperation with discovery services. This document begins with a statement of the background of the discovery services environment and proceeds to outline concerns and issues that relate to how discovery services handle content from resources such as A&I products. The recommended practices center on the rights and obligations that apply to the five categories of stakeholders: content owners, content platforms, discovery services, subscribers to content resources, and end users. A matrix is provided that illustrates the rights and obligations that apply to the specific components or activities within the discovery ecosystem relative to each stakeholder. Section 4.4 "Description and Rationale" provides a set of eighteen statements that delineate the recommended practices that apply to how discovery services would need to handle these content resources in order to satisfy all the concerns of all the stakeholder categories.

The NFAIS Recommended Practices: Discovery Services document was given careful consideration by the NISO Open Discovery Initiative Working Group (see above). ODI recognized the special considerations of A&I resources as described in this document and developed its recommended practices to apply to the broader range of content providers as they relate to discovery service providers.

Discovery: A metadata ecology for UK Education and Research

The Discovery initiative (discovery.ac.uk/) in the United Kingdom was active between 2011 and 2012 with the intent to improve discovery of resources through the improved metadata practices. The initiative included participation of a number of organizations including Jisc, Mimas, Research Libraries UK, Eduserv, Collections Trust, and Sero Consulting. One of the outcomes of the project was the development of a set of "Discovery Open Metadata Principles" that define practices which aim to improve the discoverability of resources through improved metadata creation and dissemination. The website for the project does not indicate any activity beyond the end of 2012.

> The JISC -funded Discovery programme was launched in May 2011 to create 'a metadata ecology' to support better access to vital collections data in libraries, archives and museums and facilitate new services for UK education and research.
>
> Our work will continue to the end of 2012 and is focused on advocating open data, reducing technical and licensing barriers, providing information, advice and training, and supporting exemplars.
>
> The Discovery programme takes forward the Vision of the earlier JISC and RLUK Resource Discovery Taskforce (RDTF) which has been working with partners from the libraries, archives and museums since 2010.

Other standards

Apart from the ODI recommended practice, there are few formal standards that apply generally to the realm of library resource discovery. Several protocols or standards may be used in specific aspects of the discovery ecosystem:

- **OAI-PMH** or **ResourceSync** (ANSI/NISO Z39.99-2014) to facilitate the transfer from content providers to discovery service providers. In addition to these protocols, these transfers also take place through file transfers, web harvesting, or other mechanisms agreed upon by the respective organizations involved.

- **KBART** (Knowledge Bases and Related Tools, NISO RP-9-2014) and related standards can be employed to help define the structure of the metadata transferred from content providers to discovery services.
- **Indexing and relevancy** is accomplished through entirely proprietary methods. While some of the current products make use of open source tools such as Apache SOLR, the specific implementations and tunings are not provided openly. Although some generalized information may be provided regarding how relevancy is calculated, detailed factors and methods are not.
- A variety of **application programming interfaces** (APIs) are involved in the discovery services ecosystem , but there has been little progress toward developing commonality. Discovery services need the ability to interact with resource management systems for events such as shelf status and availability for loan, for patron account features, and resource requests. These tasks are generally accomplished through a combination of library-oriented protocols such as SIP2 or NCIP and proprietary APIs specific to each resource management system. These components are assembled into connectors that can be reused across implementations of each resource management system and discovery service. When the discovery service and resource management system are developed by the same vendor, these interactions may be conducted through proprietary mechanisms.

  How an index-based discovery service interacts with discovery interfaces also lacks standardization. Services such as ProQuest Summon, EBSCO Discovery Service, Ex Libris Primo, and OCLC WorldCat Discovery Service each offer an API that can be used to connect with third-party discovery interfaces; each of these APIs differ substantially.

d. Indexes Dominated by Private Commercial Agreements

These index-based discovery services are based on pre-built indexes populated with citation data, full text, and other data elements loaded through batch processes. Much of the data that populates the indexes is supplied by content providers, especially commercial publishers, through private agreements. The universe of discoverable content through these index-based products is inherently limited by the extent to which commercial entities are willing to make these private agreements for data exchange.

One area of concern currently regards the value added from abstracting and indexing services. Libraries value these products and expect their content and metadata to be incorporated into discovery services, but some of the providers of those products balk at the notion of exposing their highly-specialized content to a generalized index.

The private nature of data exchange agreements between commercial content providers and discovery service creators evoked an interest in more transparency regarding the data used to populate these indexes and how the data are exercised within the discovery services. The NISO Open Discovery Initiative (discussed above) was formed to ensure more transparency regarding these private agreements and to promote additional participation from content providers to cooperate with discovery services providers. Opportunities for expanding the content of centralized indexes outside of private agreements may lie in the growing body of open access content and in metadata exposed as open linked data.

The current discovery ecosystem is bound by many proprietary or commercial factors:

- The body of scholarly content of interest remains largely held by publishers that offer proprietary content limited to paid subscribers
- A growing portion of scholarly content is published through open access models. This open access content is eligible for inclusion in any discovery service and for access to any researcher. The key challenge lies in identifying individual open access article or the journals in which all of the articles are open access so that they can be included in discovery indexes with full text linking enabled for all users. The NISO Access and License Indicators recommended practice (NISO RP-22-2015) published in January 2015 provides a mechanism that addresses this need.
- The number of products available in the index-based discovery service category is somewhat limited. The four major products, however, engage in vigorous competition. EBSCO currently holds the dominant market share, challenged by Ex Libris Primo, OCLC WorldCat Discovery Service, and ProQuest Summon.
- This mix of providers includes both non-profit (OCLC) and for-profit (Ex Libris, ProQuest, EBSCO) business models.
- All of the current index-based discovery services are provided exclusively through commercial business terms.
- Google Scholar provides another commercial alternative. Although Google does not charge for the service, it is closely tied to the company's other search technologies and is operated in a non-transparent manner.
- Open source products exist only for the discovery interface layer.
- No open source platforms exist for the delivery of a global discovery service.
- No open access discovery index has been created.

This environment, where many of the key components are provided only through proprietary means and commercial organizations, raises the question of whether the introduction of open access or open source alternatives could lead to lower costs or increased innovation.

e. Open Access Global Discovery Service or Index

The Index-based discovery service arena remains entirely dominated by commercial providers. EBSCO Information Services, ProQuest, OCLC, and Ex Libris have all created products limited only to those libraries that pay negotiated subscription fees. Google Scholar and Microsoft Academic Search stand as two of the few discovery services available without direct payment by libraries or end users. This section explores the feasibility of or opportunities for the possible creation of alternative indexes, such as through open access or crowdsourced projects.

The development and deployment of these services requires extensive resources, including a highly scalable technology platform; a broad program of publisher relations that negotiate and execute agreements relative to the provision of content to populate central indexes; and the development of software for interfaces, indexing, relevancy, and many other technical components that comprise these services. So far, the creation of new index-based discovery services based on open source software and an open access index has been beyond the resources of non-commercial entities to produce. The key

question on this topic would center on whether there are possibilities for lowering the threshold of complexity, expense, or business processes that have so far limited the discovery service arena to a narrow band of commercial providers.

In many other areas of library technology products, open source alternatives are available. These open source products not only provide an option of interest to many libraries to enable more customized solutions to the problem at hand, but they also provide a downward price pressure in the commercial arena. The availability of low-cost open source products can serve as a factor when commercial vendors set prices. While the direct and indirect costs of implementing open source software must be taken into consideration, commercial products set at drastically higher prices would fare less well in a competitive product arena that included both business models.

In the ILS arena, open source products, especially Koha and Evergreen, offer vigorous competition to the commercial offerings from organizations such as SirsiDynix (Symphony® and Horizon®), Innovative Interfaces (Millennium, Polaris, vtls-Virtua), The Library Corporation (Library·Solution, Carl·X), Auto-Graphics (VERSO®), OCLC (Amlib, Sunrise, OLIB, BOND), and many others. These open source ILS products are implemented in libraries through commercial support services as well as through direct efforts of libraries.

Library Services Platforms, the more recent genre of products offering more comprehensive resource management capabilities, includes the open source Kuali OLE project in addition to commercial and proprietary products, including Sierra from Innovative Interfaces, Alma from Ex Libris, WorldShare® Management Services, and Intota™ from ProQuest. The Kuali Foundation has recently made a transition to a new business model where a new commercial company was launched to take forward the development and support of its projects. The Kuali OLE project currently is expected to continue its current community-driven development, but this transition illustrates how open source projects have complex relations with commercial organizations.

In the discovery interface arena, two open source products, VuFind and Blacklight, have gained considerable adoption. These compete with the online catalogs and discovery interfaces produced by commercial companies including Enterprise from SirsiDynix, Encore from Innovative Interfaces, LS2 from the Library Corporation, among others. The open source discovery interfaces give libraries a great deal of control in the appearance of their environment and with the information resources they want to present to their users. These open source interfaces can also incorporate search results from index-based discovery services. Example implementations are Villanova University's VuFind-based interface that includes results from ProQuest Summon; Columbia's Blacklight interface that includes results from Summon; University of Chicago's new VuFind-based interface that includes results from EBSCO Discovery Service.

Platforms that support institutional repositories include more open source options than commercial. Open source platforms include DSpace and Fedora (both now under the governance of DuraSpace), Islandora, and Hydra. Commercial products include Digital Commons offered by bepress (www.bepress.com/).

It might also be of interest to see such an alternative in the discovery services arena. To understand the feasibility of such a product, it is helpful to consider the tasks and components that would need to be accomplished and what projects might currently be in place. An estimation of the resources and complexity of each task would also inform an assessment of the feasibility of creating an open source

and open access alternative based on shared personnel and financial resources contributed by libraries or aligned institutions.

An open access discovery service could be delivered through one of the open source discovery interfaces. There are at least two products currently available that would be able to take advantage of an open access discovery index if it were able to offer APIs similar to those available from the commercial index-based discovery services. Some enhancements of these interfaces may need to be accomplished to strengthen their architecture and improve functionality, which is already taking place as new implementations are made in local discovery environments.

The creation of an open access alternative in the discovery services arena would depend on the feasibility of a grant-funded or community-sourced project able to create and maintain each of the components that comprise a discovery service. As already noted, open source discovery interfaces, such as VuFind, Blacklight, and the eXtensible Catalog toolkit are currently available to provide the end-user interface. The creation of a central index poses a much greater challenge. Although there are multiple possible approaches, such a service would include a variety of tasks such as the creation of a technical platform to manage the index, processes to gain access to content for indexing from publishers, and the processes to maintain the currency of the index.

Local indexing scale

It is important to differentiate the indexing usually associated with a local discovery interface product versus those provided through discovery services based on a comprehensive central index. Although discovery interfaces often maintain a local index, it is usually populated by records representing collections of finite scope. The most common scenario involves populating the local index with MARC records extracted from the ILS implementations of a library or consortium. Some may also represent local collections maintained by the library or specific remote repositories of interest. In the public library arena, the index of a discovery interface is often populated with the records of large e-book collections, such as has been accomplished by the Marmot Library Network in Colorado using VuFind as a component in its statewide e-book lending platform.

The index of a discovery interface has not been used to load article-level collections of the scale that approaches what is provided by the index-based discovery services. These local indexes may be populated by many millions or tens of millions of resource records, but do not approach the billions of content items generally understood to be represented in commercial discovery services.

The scale and scope of a central discovery index should also be distinguished from the knowledge bases associated with link resolvers and electronic resource management systems. These knowledge bases contain title-level metadata, linking syntax rules, publication ranges and other data. These knowledge bases that aim to describe the body of library-oriented electronic resources manage in the order of 1-2 million data elements. [See: Breeding, Marshall. (2012).] Commercially-produced knowledge bases include those from Ex Libris, ProQuest, OCLC, and EBSCO Information Services. The Global Open Knowledge Base (gokb.org/) aims to produce an open access knowledge base created through collaborative participation of libraries, supported through funding of the Andrew W. Mellon Foundation. GOKb v4.-.4 Public Preview is currently available. A recent login into this system shows coverage of just over 2,000 publisher organizations, 23,000 titles, and 254 content packages. While these numbers represent important progress, they remain considerably below the quantity of data managed in the commercial knowledge bases. The Global Open Knowledge Base provides an opportunity to observe

whether a community-sourced open access project can be created to operate at a comparable level as the commercial alternatives and result in introducing new innovations, improved data quality, lower costs, or other desirable outcomes.

Knowledge bases have also become less of a product in themselves and a more of a commodity that functions as a component within broader technology platforms. As link resolution and electronic resource management become folded into the functionality of library services platforms such as Alma, OCLC WorldShare Management Services, ProQuest Intota, and Kuali OLE, they remain important but are diminished as a separate point of innovation or distinction. Questions of innovation versus commoditization may also apply to the creation of an open access discovery service that involves the creation and upkeep of an index managing billions of items of content versus at most a few million as seen in the genre of e-resource knowledge bases.

Central index

Index-based discovery services include an index generated from records and content items that aim to represent the totality of the body of scholarly content and other materials of interest to libraries. These massive indexes aim to cover a specific subset of library-oriented content, resources, and services available on the Web. The provision of a central discovery index includes both the gathering of the content items that will be used to generate the index and the provision of a technology platform to manage the index and deliver its functionality.

One major challenge for any proposed open access implementation would involve the implementation of a technical platform capable of indexing 1-2 billion content items, including a high portion of very large full-text records.

Some of the needed components for such a technical platform might include:

- Storage capacity for multiple replications of the index
- Highly scalable indexing, search, and retrieval technology
- Software utilities for processing records received from content providers for indexing. (Processing tasks might include normalization, error detection and correction, and staging into index loading modules.)
- Software layer to provide API to respond to discovery interface requests
- Multi-tenant software to provide access according to institutional or individual identities

The creation of a technical platform on par with the commercial discovery services would involve the development of software to load, index, and perform search and retrieval operations. While there are many open source components available that could form the basic infrastructure, considerable development would be required to provide a functional system.

While it is difficult to estimate the level of resources needed for the technical development of such an open access discovery service, data from the efforts of the commercial providers can help establish the scale. EBSCO, for example, reports: “We have more than 300 technical staff working on discovery alone….” [discovery.ebsco.com/pulse/article/welcome-to-discovery-pulse]. Other vendors have not made public statements describing the personnel and technical resources allocated to their discovery services.

Issues regarding proprietary content

The construction of an open access discovery index must also address the intellectual properties related to the content resources indexed. Much of the content of interest remains under the proprietary ownership of publishers. The commercial discovery services include proprietary content, usually under the provision that it not be exposed publicly, be used only in the construction of an index, and that any display or linking of proprietary content be provided only to mutual subscribers. Proprietary content includes such items as the full text of non-open access articles, and abstracts created or terms assigned by commercial A&I services. While a more ideal future environment might be based on mostly open access content and a robust universe of open linked data, the current reality requires discovery of content with many layers of restrictions.

Citations and some levels of metadata have few if any access restrictions and can be treated more freely. It would be possible to develop an open access discovery index that included only citations to articles, but such an approach would likely be considered inferior to the commercial products currently available that provide extended retrievability through indexing of proprietary content. For non-open access journal articles, their full text would fit in the category of proprietary content. Subject terms, abstracts, and other elements produced through proprietary abstracting and indexing services could also not be provided in an open access model.

A more complete solution would require receiving full text and other proprietary content from providers under similar provisions as commercial discovery service providers. This model would require stringent controls—both legal and technical—to ensure that proprietary content not be publicly exposed.

The retrieval and ingestion of content for the population of the central index could be accomplished either through a managed content transfer processes similar to how the commercial index-based discovery services receive data from content providers or through a harvesting method, such as that used by Google Scholar. Either method would require cooperation of the publishers to gain access to content that is otherwise restricted to subscribers for the purpose of indexing.

Incorporating proprietary content would have deep implications on the creation of an open access discovery index. It might limit possibilities for broadly distributed workflows among libraries or other institutions participating in its development. It would be challenging to convince publishers to provide their proprietary content assets to an open access discovery service managed by hundreds or thousands of library participants. Such an open access discovery service might be managed on behalf of the library community that would be less diffuse and able to negotiate access to proprietary content.

There is already a non-profit discovery service operated by an organization that is owned and governed by the library community. WorldCat Discovery Service, developed and managed by OCLC, fits into this category. Yet, WorldCat Discovery Service does not seem any closer to an open access business model relative to the index-based discovery services offered by for-profit companies. Access to WorldCat Discovery Service is limited to libraries with paid subscriptions. In pursuit of more open options for an index-based discovery service, there may be some possible arrangement with OCLC, but it would involve a substantial deviation from their current business practices.

Publisher relations

Each of the commercial discovery services is supported by a team or workgroup that deals with the content providers supplying materials to populate the index. Any project to create an open access

discovery interface would also need an extensive network of participants to initiate discussions and coordinate content contributions from publishers. Given the thousands of publishers offering content of interest, the number of personnel resources required for this set of tasks would be considerable. But since libraries deal with these same publishers in acquiring and managing their subscriptions, they are well positioned to perform this set of tasks.

Automated content contribution

It could also be postulated that the population of an open access discovery index could be performed through automated processes. The service could rely on retrieving content from the servers of publishers using protocols such as OAI-PMH or ResourceSync. This method would require that publishers be willing to expose their entire set of content offerings and provide the technical environment to respond to requests from open access discovery service providers. While this approach is routinely used for institutional and disciplinary repositories of open access content, it has not been proven as a comprehensive tool for harvesting proprietary content from publishers.

The NISO Open Discovery Initiative did not define a standardized mechanism for the transfer of content from discovery service providers to discovery service creators. Rather the group suggests that content providers make use of existing standards to conduct the transfer of content and metadata:

> The ODI recommends that the transfer of data from content providers to discovery service providers should make use of existing standards where applicable. Some of the standards and protocols most directly applicable include the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and KBART. ResourceSync also has strong potential as a mechanism for data transfer for discovery services once it is published by NISO and software tools become available for its implementation.
>
> *Open Discovery Initiative: Promoting Transparency in Discovery* (NISO RP-19-2014)

It would be exceptionally optimistic to expect that an open access discovery index could be populated by a mostly automated set of procedures any more than this has been implemented for the commercial discovery service providers. It may be more feasible to make arrangements for web-based harvesting that bypasses the paywalls of publisher servers following the model of Google Scholar.

Open access discovery index without a platform

One possible scenario might involve the development of an open access discovery index independent of a platform from which it would be deployed for production use. Libraries or other organizations interested in making use of the index would provide their own platform to load the content of the index, providing their own software stack and hardware platform.

The open access discovery repository would consist of the raw citation records, full-text articles, and other resources that would then be available for indexing. Many complications would apply to this scenario, especially related to the handling of proprietary content to prevent unauthorized access.

Such a content repository in support of an open access discovery index would need to provide a centralized staging or storage facility but without also developing the platform required for indexing and query responses. Libraries or other organizations interested in developing an index based discovery service would provide their own search platforms, relying on the central content repository to populate its indexes. If such a resource were created, interesting questions might apply, such as whether the commercial index-based discovery services would also have access.

Viability of open access alternative

This review of the components of an index-based discovery service highlights the enormous level of resources required to create and maintain them. The creation of an open access discovery index would require the allocation of capital, personnel, and technical resources at least at the level of what any of the commercial providers has devoted to their projects.

The index-based discovery arena currently has four strong commercial competitors. The number of viable alternatives for resource management systems for academic libraries may be even narrower. The interest and motivation for the library community to pursue an open access and open source alternative in the index-based discovery service arena would need to be driven by high levels of dissatisfaction with the current products, pricing considered unreasonable, lack of innovation, or other similar factors.

2. Integration of Discovery Services with Resource Management Systems

An important issue in the current environment relates to the degree of independence between resource management systems and discovery services. To what extent can these products be selected and implemented from different providers? Libraries may prefer a discovery service based on its functionality and content coverage and may prefer a resource management system from another vendor based on another set of distinct requirements. Many libraries need the ability to set discovery and management strategies independently and expect these systems to have mutual interoperability.

Discovery services include mechanisms for access-appropriate services related to local content items. Once common scenario involves the treatment of a library's print collection. The discovery service must be able to provide the functionality of the online catalog module of an ILS, including displaying a status of an item's availability and its shelf location, and to be able to provide services such as placing a hold request. Discovery services also interact with the institution's integrated library system as they assume responsibility for patron account features that were previously provided through the online catalog module. These ILS integration capabilities are increasingly expected in a discovery interface as libraries increasingly retire their online catalog modules. The integration between a discovery service and an ILS has been addressed through the ILS-DI initiative of the Digital Library Federation. [See Breeding, Marshall. (2008).] This integration may be accomplished through a combination of established protocols, such as SIP2, NCIP, and Z39.50, but also involves tasks not directly addressed by these standards. To enable full interoperability, additional APIs or other mechanisms must be enabled in the ILS.

The APIs and other mechanisms to enable integration of an ILS with a discovery service are under the control of the provider of the ILS. The discovery service provider likewise controls the APIs needed to integrate with its product. The ability for a library to use any given discovery service as their comprehensive patron-facing discovery interface depends on mutual access to both sets of APIs. In some cases, discovery service and ILS providers have business partnerships that explicitly deliver such integrations. In other cases, libraries as customers of these products have access to the APIs and can perform their own integration. Cases remain, however, where ILS providers and discovery service providers do not cooperate, making some scenarios of integration difficult or impossible.

Libraries benefit from some of the efficiencies possible when the data and business logic of the resource management system is directly available to the discovery service. But any unbreakable coupling between specific discovery services and resource management platforms imposes concerns for libraries. Libraries often have an interest in the ability to use their preferred discovery service regardless of the resource management platform in use. Those that have implemented a discovery service that meets their requirements may not want to be forced to change discovery services if they replace an outdated ILS with a library services platform offered by a vendor other than the one that provided their discovery service.

Many vendors of resource management systems (including both integrated library systems and library services platforms) also offer discovery services. Table 1 provides some examples.

Table 1: Vendor offerings for discovery services and resource management products

Provider	Discovery Service	Resource Management
Ex Libris	Primo	Alma, Aleph, Voyager, Verde
ProQuest	Summon	Intota, 360 Suite
OCLC	WorldCat Discovery Service	WorldShare Management Services
EBSCO	EBSCO Discovery Service	None (some ERM tools)

In the current field of competitors, the only provider of index-based discovery services that does not offer its own resource management product is EBSCO Information Services.

The open source Kuali OLE resource management system does not include its own discovery service. It has been designed to expose APIs to enable libraries to implement the discovery service of their choice.

The discovery services and resource management products from a given vendor will naturally have a built-in affinity. These products are offered as a pre-integrated suite that may use APIs and proprietary technical mechanisms. These product suites are also based on common concepts of design and technical architecture. Content components, such as e-resource knowledge bases, also unite these products. Tight coupling between discovery services and library services platforms offers efficiencies. But tight coupling that precludes integrating discovery and resource management products from separate vendors eliminates choice.

Some providers of discovery services do not offer resource management services and have a strong interest in being able to integrate with all of the major resource management systems. Vendors that offer both resource management systems and discovery products naturally prefer that libraries implement these products together. Many libraries opt for these combinations out of cost incentives, smoother end-user workflows, and simplified implementation and support. Other libraries may prefer a mix-and-match approach. Vendors that offer only discovery services naturally want the ability to integrate with the full field of resource management products.

Many libraries may also prefer to assemble their own discovery environment based on an open source tool, such as VuFind or Blacklight, programmed to interact with their own choices of resource management systems, index-based discovery services, or other content repositories.

In addition to concerns in how discovery indexes are populated, an area that would benefit from additional transparency and mandates for cooperation lies in the interoperability of discovery systems with resource management systems and other relevant components in the information infrastructure within a library, that of its higher-level organization, or with the platforms of external providers.

3. Linked Data

The current model of index-based discovery seems likely to persist for the indefinite future. These platforms will become increasingly powerful tools for providing access to library collections, especially if their ecosystem evolves toward universal participation. Yet, this model will not remain unchallenged indefinitely. The current momentum seen with open linked data will likely lead at a minimum to extensions of the index-based model or hybrid systems, with a longer-term possibility of discovery services based entirely on linked data rather than harvested citations and full text.

One of the main areas of interest in the broader realm of information systems lies in the realm of semantic web technologies or open linked data. This approach to exposing and exploring resources has become increasingly adopted outside of the library environment. Of all the many developments in the information arena, open linked data has the greatest potential impact on discovery services. The use of open linked data is rapidly moving from the experimental and prototype phase into an operational technology that warrants further attention in the discovery arena.

Library resource discovery services are natural beneficiaries to the expanding universe of open linked data. Index based discovery services continue to push the limits of what can be accomplished with keyword-oriented search and retrieval technologies. Yet this approach comes with some inherent limits, given that it is based on a monolithic index that must contain all the content and relationships for an extremely expansive body of information and knowledge. The realm of open linked data provides opportunities to leverage content and relationships outside of what can be bound within a discovery index. Exposed linked data also serves as a source of content that can be harvested and indexed by the current model of index-based discovery services.

a. Tools and Technologies

Open linked data technologies have not yet been proven to be scalable, reliable, and able to sustain high-performance operational systems. The current phase of discovery based on central indexes built from metadata and full text provided by content providers in bulk has demonstrated the minimum scale at which discovery services have to operate to be considered useful. These indexes include representations of over 1 billion items and sustain significant operational transaction loads. Services based on open linked data will need to sustain similar scale and performance thresholds.

Open source tools, such as Apache SOLR, are available that provide industrial strength performance for traditional index-based search and have been widely adopted in library-oriented products. The availably of these components has accelerated the development of the current slate of index-based discovery services. The identification or creation of similar infrastructure components or toolkits that can be applied to library-oriented discovery products based on open linked data would likewise accelerate progress in developing new services.

There may be possibilities for integrating tools and resources based on open linked data with index-based discovery services to create hybrid systems that take advantage of the strengths of both discovery models.

b. Exposed Content Resources

In the same way that not all content, structure, and relationships can be bound within a discovery index, we are also at a time where only a limited portion of the universe of content of interest to libraries has been exposed as open linked data, although there does appear to be a rapid increase in scholarly resources being exposed as linked data. The current concern in the discovery arena centers on content providers providing private access to their resources only to discovery services providers. The next phase of open discovery might express mandates that content providers expose metadata as open linked data. If scholarly content were exposed as open linked data, significant new innovations could be enabled. A shift to open linked data could also open the discovery services arena to allow for the development of services beyond the current small band of commercial providers based central indexes built through private arrangements with content providers.

Discovery services will inevitably have to become more oriented to linked data. BIBFRAME is positioned to eventually displace MARC as the substrate for carrying bibliographic data. While the timeframe in which resource management systems transition from MARC to BIBFRAME will likely be many years, discovery services will likewise need to make that same transition. These services will need to optimize how they index bibliographic data based on linked data oriented structures such as BIBFRAME. They could also enhance the content of their indexes through following and indexing resources referenced in links that may not have been possible with MARC records. Unique identifiers and URIs derived from BIBFRAME structures may enable improvements in access and linking.

There realm of open linked data of interest to library discovery continues to increase. Collections under library control and not constrained by commercial agreements or copyright limitations are increasingly being exposed as open linked data. Currently, organizations with the technical capabilities are active in exposing their collections. As the mechanisms for converting and exposing collections as open linked data become more available and with lower thresholds of difficulty, the quantities of exposed content will increase dramatically.

Many national and international projects heavily rely on open linked data as part of their architecture. These projects include:

- VIAF: Virtual International Authority File (viaf.org/) [Hosted and managed by OCLC.]
- Digital Public Library of America (dp.la/)
- Europeana (pro.europeana.eu/linked-open-data)
- Library of Congress: subjects and authorities. (id.loc.gov/)
- National Library of Sweden: LIBRIS Swedish Union Catalogue (librisbloggen.kb.se/2008/12/03/libris-available-as-linked-data/)
- British Library – Free metadata services (www.bl.uk/bibliographic/datafree.html)
- German National Library (www.dnb.de/EN/lds.html)
- Biblioteca Nacional de España [Datos enlazados en la BNE] (datos.bne.es/)
- National Library of The Netherlands – Research stage (researchkb.wordpress.com/2014/08/26/linked-open-data-at-the-national-library-of-the-netherlands/)

- Elsevier: Linked Data Repository (LDR) (data-sandbox.elsevier.com/documentation/index.html)

Datahub is a platform that aggregates open linked data sets. There are currently 18 data sets in the Library Linked Data group (datahub.io/group/lld).

Categories of content that lend themselves to exposure as open linked data include:

- **Bibliographic databases** from individual libraries, consortia, and regional or national libraries can be used to expose the bibliographic records that describe their print collections and other materials managed within their ILS. How overlap of the same records from individual libraries versus collective catalogs is handled in the universe of open linked data will be a complex issue.
- **Digital collections**, **finding aids** for special collections, institutional repositories, and other sets of resources created by individual libraries can gain enhanced availability when exposed as open linked data. Techniques such as enhanced mark-up to include schema.org encodings can also enhance discoverability on the open web.
- **Content of library websites**, subject guides, and other library created web content can likewise gain broader exposure through schema.org or other techniques to expose this content through semantic web technologies or as linked data.

Key questions include whether a critical mass of the metadata describing the body of scholarly literature currently held by proprietary publishing arena will be exposed as open linked data.

Zepheira, a consulting firm that has been one of the major players in the linked data arena for libraries, has launched an initiative it calls **Libhub** (libhub.org). The initiative centers on providing better exposure of library resources via BIBFRAME and other linked data structures. The stated goal of Libhub is to "raise the visibility of Libraries on the Web by actively exploring the promise of BIBFRAME and Linked Data." The initial activities of the program involve harvesting bibliographic records in MARC format from a group of participating libraries, converting those records into BIBFRAME using Zepheria's tools, publishing the BIBFRAME records, and then analyzing whether discoverability or visibility of those resources in their owing libraries increases as a result.

Another interesting project, called Serendipity, has been launched at the La Universidad Técnica Particular de Loja to explore course-related resources through a linked data browser. This project does not seem to be directly related to the library, but comes out of the Computer Science Research Institute - I2C2. The project is based on j4loxa, a general linked data exploration tool (j4loxa.com/). This project provides an example of using linked data for discovery in a very limited content domain. Developers characterize the tool as providing Linked OpenCourseWare Data Faceted Search:

> Consequently, we argue that these advances are a possible means of supporting interoperability, accessibility and reusability of the data types like the Open Educational Resources and OpenCourseWare (OER and OCW).
> (http://serendipity.utpl.edu.ec/about.html)

4. Gap Analysis

This section aims to identify some of the features or concepts that may not be fully realized in the current generation of discovery services. These features include those that can be implemented by incremental enhancements to the current products. Others may involve systemic changes relative to the current ecosystem of content providers and discovery services. What are some of the gaps between the functionality delivered by the current generation of discovery services and that expected by library users and librarians?

a. Coverage of Relevant Resources

The central indexes of the major discovery services continue to expand, working toward a goal of comprehensive representation of the content resources of interest to libraries. Yet, omissions in coverage remain. Understanding the exact extent of what is covered or not remains problematic. While discovery service providers generally make available resource lists of content indexed, these often do not provide the level of detail needed to enable librarians to perform detailed queries or systematic analysis of content coverage. No third-party tool is currently available that enables libraries to compare content coverage of discovery services. Libraries assessing or analyzing the coverage of discovery services would benefit from tools created and maintained from independent and trusted third-party individuals or organizations.

Primary publisher content

Coverage of the mainstream scholarly content from primary publishers based in the United States and Europe is generally strong, with gaps narrowing. Primary publishers are well motivated to ensure that their content is well represented in any of the discovery environments used by their subscribing libraries. Common issues relate to the delay between availability of new material by publishers and the time in which it becomes available in the indexes of the discovery services. Significant room for growth remains through increased inclusion of scholarly and cultural materials from diverse international regions and languages.

Abstracting and indexing services

Many A&I providers continue not to contribute their proprietary content to discovery services. The NFAIS *Recommend Practice: Discovery Services* articulates many of the issues and concerns that discovery services must address relative to handling proprietary content, such as that related to A&I products. Many issues remain unsettled regarding how discovery services handle A&I data related to indexing and treatment of their value-added proprietary content, which continue to impact the participation of these vendors with index-based discovery services.

EBSCO Information Services is involved in multiple activities within the industry. The company is a major provider of content products, produces a variety of A&I products, and offers EBSCO Discovery Service. EBSCO has issued an Open Metadata Sharing Policy (www.ebscohost.com/metadata-sharing-policy) that offers content from its non-A&I products to other discovery services. The offer includes reciprocal terms that require that ILS vendors should also enable interoperability from their resource management products with EBSCO Discovery Service. EBSCO Discovery Service emphasizes A&I content in its search architecture and includes its own A&I content as well as many non-EBSCO owned A&I resources.

Ex Libris and EBSCO currently remain deadlocked regarding an agreement that would allow EBSCO's A&I products to be indexed by Primo.

Beginning in 2014, ProQuest began providing content from its A&I products to Ex Libris and OCLC. A press release dated September 29, 2014 announced that the indexing of major ProQuest databases, including ProQuest Central, ProQuest Dissertations and Theses Global, ABI/INFORM, Black Studies Center, Early English Books Online, Periodicals Archive Online, Periodicals Index Online, and approximately 50 abstracting and indexing databases are now indexed in Primo Central.

The general participation of A&I resources in the discovery services arena remains moderate to weak. There have been some isolated examples of new engagements by A&I products with discovery services. The non-participation or incomplete contribution of A&I content to index-based discovery services remains one of the largest gaps in the current set of products and a topic of considerable controversy. It seems apparent that further work needs to be done in the library and vendor communities to resolve these issues in ways that will lead to the inclusion of A&I content in all of the discovery service as libraries expect.

b. Internationalization and Multi-Lingual Coverage

Coverage of bibliographic resources from diverse international sources is growing. OCLC has been especially active in gathering bibliographic records from the catalogs of national libraries globally, spanning many languages and scripts. Many of these international bibliographic collections may not be available to other discovery services.

Coverage of article-level scholarly resources, primary research resources, and other material in non-English languages is likewise improving, but is far from universal. The content represented in discovery indexes is becoming increasingly heterogeneous by language, which introduces challenges in search and retrieval.

As discovery services are implemented in international regions, agreements are often formed to include major resources from that area in their indexes. These arrangements make the discovery services more viable in these regions, but they also provide the benefit of providing expanded exposure of those materials globally.

Although the coverage of materials published in specific countries or global regions in other languages has improved, it has not advanced to the point where libraries in those regions can depend on discovery services for local materials. Libraries in Asia, for example, might benefit from the current discovery services to provide better access to Western materials, but these tools do not currently provide comprehensive access to the scholarly materials in their own region. Many countries and regions, often through their national library, continue to offer and develop their own catalogs or discovery services of local materials that may not be well represented in the global discovery products. Considerable opportunity remains for improved representation of scholarship and research spanning non-Western and non-English cultures and regions.

Technologies and techniques to query concepts or terms across multiple languages remain limited. Cross-language searching remains fertile ground for future development in discovery services. The success of these products in other international regions is a factor of how well the scholarly literature of that country or region in the local languages and scripts is available in their indexes and supported in the

interface and query features. As technologies that support multi-language searching by concepts mature, they will be a great asset to the discovery services arena.

c. Coverage of Open Access Materials

The major discovery services are beginning to provide better coverage of open access materials, but room for improvement remains. Each of the discovery services includes commercially published open access titles, materials from major disciplinary open access servers such as aXchive.org, and can tap into open access materials through centralized services such as OAIster. As the scholarly publishing arena shifts toward greater proportions of open access, discovery service providers are well motivated to adjust their content coverage accordingly.

The recently published NISO recommended practice on *Access License and Indicators* (NISO RP-22-2015) suggests use of new metadata that can be leveraged to improve the representation of open access materials in discovery service. Once the ALI-recommended metadata is implemented by publishers that include open access materials, discovery services will have an improved ability to identify open access materials as they are indexed and to enable access independently of the library's subscriptions.

The challenge is how to expose open access materials from a variety of sources:

- Discipline-oriented repositories
- Institutional repositories
- Discretely published open access journals (not necessarily represented in commercial packages)
- Open access materials within commercial products to which the library does not subscribe

The issues that arise in providing access to open access content are not necessarily that complex. As libraries identify open access materials not previously covered, they can add these to their local indexed materials or prompt the discovery service provider to index them globally. The inclusion of open access materials does not involve many of the complicated business negotiations between publishers and discovery service providers that apply to proprietary content.

d. Precision and Known-Item Searching

Advanced and precision searching continue to be areas of interest for discovery services. The complex and non-intuitive user interfaces of the earlier generation of online catalogs eventually led to the emergence of next-generation catalogs, or discovery interfaces. The newer set of resource management systems that manage a wide range of materials beyond those of the ILS do not supply a catalog scoped to the print collection and rely instead on a discovery interfaces for all patron access. As much as online catalogs were not fully appreciated by some library users, they excelled at providing precise methods for interacting with a library's local collections. These traditional catalog interfaces enabled patrons to browse through collections based on name or subject authority databases, to virtually browse items as they would be shelved based on call number indexes, and to perform advanced Boolean queries. Many of these techniques are more appreciated by librarians then by library users. Yet many students and researchers likewise require the ability to perform precise queries and expect comprehensive and orderly results.

Many of these search techniques are based on structures in resource management systems that do not exist in the broader universe of electronic resources. The authority work performed for monographs as they are cataloged by libraries has never been applied at the article level of electronic resources. Even though A&I services apply structured metadata, they tend to be based on discipline specific ontologies. Names of creators are especially inconsistent between article metadata and monographic cataloging. Conventions for citations for scholarly articles, for example, tend to be based only on initials, whereas LC Name authorities use full representation of names, in inverted form. Projects such as ORCID (orcid.org/) or ResearcherID (www.researcherid.com/) have potential for bringing all the items by the same author together within a discovery environment through use of an author identifier. However, it seems that comprehensive application of these unique identifiers across all authors and their current and retrospective works remains quite a distant prospect.

One additional challenge lies in the ability of discovery services to find known items. Especially when searching for resources with one-word titles or common words, such as *Nature* or *Time*, relevancy-based retrieval may not always return the expected results. Each of the discovery services has improved its handling of known-item searching, but this continues as a point of criticism of performance.

e. Relevancy Rankings

The way in which discovery services order search results is critical. Given the massive body of material represented in their indexes, it is essential for the items that best match the query to appear near the top of the results list. Users may not have the patience to work through many pages of results; those without extensive knowledge of the discipline may select resources to use for their research that may omit items of greater importance if they happen to fall deeper in the list. Any anomalies or poor relevance ordering will be readily apparent to librarians and experts in a specific discipline.

Improving relevancy ranking has been a high priority for the developers of discovery services. Each of the vendors has developed tools and technologies to improve their relevancy performance. Ex Libris, for example, has developed what it calls ScholarRank to inform relevance ranking factored into associations derived from its bX Recommender service. EBSCO Discovery Service describes its general approach to relevance ranking in a public document (www.ebscohost.com/discovery/technology/relevance-ranking).

Relevance ranking remains one of the key issues that impedes support of librarians for these products. Many librarians characterize the performance of discovery services as unpredictable and erratic in the delivery of search results (as reflected in survey responses and general interactions with public service librarians). Further improvements are needed in this aspect of discovery services.

How relevancy functions within each of the discovery services remains in the proprietary realm and is considered one of the main competitive features. (Details regarding relevancy was considered out of scope of the NISO Open Discovery Initiative.) Expectations for transparency in how discovery services calculate relevancy could be a positive factor in improving the performance and the acceptance of these products.

f. Enhanced Discoverability through Non-Textual Associations

Apart from the value-added content from A&I products, discovery services aim to provide enhanced discovery beyond keyword matching through other means. They may, for example, be able to enhance indexing and retrieval functionality to perform some level of query enhancement and facilitate the

retrieval of relevant materials even when the user does not enter query terms that align with the vocabulary used in the full text of the articles. Clustering technologies may be able to produce facets based on the content of articles retrieved to guide the searcher toward the ones that match their interests. Other technologies employed in the current generation of products include exploiting various types of use data to improve retrievability and relevancy. Examples include the proprietary bX technology that makes use of associations reflected in link resolver logs to identify works that may be related even if they do not include the same keywords. This technique builds on the likelihood that articles may be related in content if they are selected by the same researcher within a particular search session.

Though progress has been made, discovery services still have much room for improvement in the way that they identify search candidates and order search results to prioritize the most meaningful items.

g. Mechanisms for Linking to Resources

One of the most critical operations of a discovery service lies in how it provides access to articles or other items of content when selected by the user. The OpenURL standard (ANSI/NISO Z39.88) provides a mechanism for context-sensitive linking designed to provide access to the full text of an item or other services based on its availability within the library's subscriptions and other factors. The classic implementation of OpenURL resolvers involves a button that launches a menu allowing the user to select a link to full text, initiate an interlibrary loan or document delivery request, or another process that might be available to provide access to that item of content. These link resolver menus are not necessarily easily understood by users. The services available may not include direct presentation of full text; instead the link may be to the journal in which an article was published, rather than to the specific article itself.

One of the questions that arises regarding the ongoing role of OpenURL in discovery services is whether it should become more of a transparent mechanism and less of something that presents its own interface to end users. Some discovery services have implemented techniques that avoid the OpenURL menu when the full text is actually available. Links to documents are pre-calculated, making it possible to simply present an icon or link that directly launches the document. These more intelligent linking strategies have the potential to result in fewer dead links and less user frustration in accessing the full text of documents.

In some cases multiple copies of a resource may be present in a library's collection given the overlap in content packages. Discovery services can handle these linking scenarios in different ways, including imposing the choice on the end user or allowing the library to set preferences regarding the priority of packages or publishers. This question of "fair linking" (i.e., whose version of the content gets provided when more than one is available) was addressed by the NISO Open Discovery Initiative in its recommendations. But as more "smart linking" takes place at a deeper level within discovery services, it may be difficult to discern how full text is presented to users in cases of overlapping content.

h. Learning Management Systems

Discovery services are optimized to provide access to resources through an interface provided by the library. It is also critical for the content and functionality of these services to be available through the interfaces of other services that are part of the natural environment of the user. Students in most colleges and universities, for example, must interact with learning management systems in the routine

performance of their work for each course. Any ability to provide functionality through such interfaces can expand the value of the discovery service.

Learning management systems are one of the key interfaces of interest to library resource discovery services. Students and instructors interact with these products constantly in the teaching and taking of courses. The ability for instructors to identify reading materials held by the library for a course represents a significant opportunity. Reserves have diminished as a routine service of the library in favor of instructors or their teaching assistants adding electronic documents to course pages. Interoperability with the library's discovery service provides a convenient way to engage instructors with resources from the library's subscribed collections.

In the UK, a genre of reading list management tools has become popular, including Talis Aspire and the open source Rebus:list supported by PTFS Europe. These products likewise benefit from interoperability with library content via discovery services.

Areas of future development in this area might include the exploration of the APIs that would benefit the interoperability between discovery services and learning management systems or other products within the campus enterprise that depend on library resources.

The Learning Tools Interoperability (LTI) framework is well established in the learning management system arena. Plug-ins or other tools that follow LTI can be implemented in any of the learning management products, saving the work of creating separate interoperability layers as is often the case when integrating library services with non-library applications.

Some projects have been recently announced in this area. Ex Libris has announced (January 28, 2015) that it will create a reading list product, though details of the planned approach were not mentioned. EBSCO released a product called Curriculum Builder that allows instructors to create reading lists from library resources through the interface of the learning management system, using the LTI framework to tap into the content and functionality in EBSCO Discovery Interface.

5. Opportunities for Future Enhancements in Discovery Services

The genre of discovery services has been developing since the initial products were introduced in 2009. Each of the products is stable and mature, with ever increasing sophistication of functionality and improved content coverage. Each of the current products has been developed by very large organizations with deep financial and development capacity. Yet, ample room remains for improvement. We can anticipate that each product will continue to be enhanced to add new functionality and capabilities in response to requests from libraries and to improve their commercially competitive position. There are other improvements and enhancements that may not arise in this way, but may relate to collective or cooperative needs that libraries express. This section explores some of these advancements in discovery.

a. Expectations Regarding Application Programming Interfaces

Each of the discovery services exposes a set of APIs to provide programmatic access to its functional capabilities for external systems. These APIs allow the discovery services to connect with resource management systems for statuses and requests related to physical resources, with third-party discovery interfaces, or with learning management systems. In the current environment, each discovery service defines its APIs independently.

Many implementation scenarios depend on the APIs exposed by discovery services. It may not always be entirely clear what APIs are available and what restrictions may apply. Libraries should expect transparence regarding APIs as they evaluate, implement, and operate any of the discovery services. Libraries interested in working with their discovery services in more complex implementation scenarios benefit from details such as:

- Complete disclosure of what APIs are exposed
- Clarity regarding the business rules regarding access to APIs:
  - Access only by current customers
  - Access by third-party developers
  - Access by competing developers
  - Access by the general public
- Business models for API access:
  - Part of the base product
  - Additional subscription cost?
  - Metered access
- A consistent set of APIs available across diverse discovery services
- Establishment of a minimum set of APIs that should be expected to be available in any discovery service.

Some of the new-generation library services platforms do not offer traditional online catalog modules. For these implementation scenarios, among others, it is important for libraries to be able to work with

discovery services to deliver the complete features set of patron-facing functionality through third-party discovery services.

b. Expanding API Ecosystem

Libraries increasingly expect the technology-based products that they acquire to not be closed systems, but rather to offer APIs or other mechanisms that enable them to build additional services or variant features beyond that provided in the base product, to dynamically exchange data with external systems, or to programmatically communicate with other systems to enable integrated interfaces for end users or to eliminate duplicative data entry by library personnel. In the current phase of technology evolution in libraries, these APIs are increasingly expected to be delivered as RESTful services. In most cases, APIs are defined by each technology provider. There is an increasing cohesion among the APIs offered by the different products or services from any given vendor, but not necessarily among the systems of competing vendors.

Given the interest in developing more APIs to enable interoperability and extensibility for each product, there is a window of opportunity for a set of cross-vendor APIs to be defined within each of the areas of intersection among products. Such an ecosystem of interoperable APIs might not be codified as standards, but instead as recommended practices that can be validated with compliance assessment. Some examples might include:

- Interactions between discovery indexes and discovery interfaces
  - Transfer a query from a discovery interface to the discovery service
  - Take advantage of facets, limiters, and other advanced search methods
- Interactions between discovery indexes and resource management systems
  - Integrated library Systems
  - Library Services Platforms
- E-book lending platforms
  - Provision of metadata for discovery
  - Dynamic availability of e-book items
  - Check-out an e-book, including any interactions with the DRM enforced by the e-book content management platform
  - Ability to integrate e-books charged within the patron profile managed by the resource management system
  - Ability to return and release e-books

c. Social Features – Communities of Collaboration

One area of opportunity for further development lies in the increased social interactivity with the realm of discovery services. Rather than simply providing search and retrieval functions against a body of content, many libraries are interested in enabling individuals to interact with these collections in a variety of ways. Collaborative communities of scholars might be able to lend their expertise within a subject discipline to provide additional points of access, or to express relationships among materials

beyond the possibility of library-based cataloging or commercial abstracting and indexing services. (See: Adams, Jonathan. (2012).)

Each of the developers of discovery services could build such capabilities within their own environments. Extending these concepts to facilitate collaborative interactive networks of scholars across discovery services would increase the possibility of achieving the critical mass necessary to be effective. Discovery services would also benefit from the ability to interoperate with existing social networks of scholars external to those defined or provided by the library community, but that might exist in external organizations. Opportunities to enable such social interactions would depend on standardized mechanisms that enable interoperability between the ecosystems of discovery services and those of external social networks. Points of interoperability would include authentication, identity management, and standard transactions for the interchange of social actions relative to the metadata associated with content items. One of the issues might lie in propagating end-user actions on a content item, such as adding a descriptive tag or creating a link of association with another content item across multiple discovery environments.

Some of the current products offer features oriented to social engagement. These features are especially prominent in the discovery interfaces oriented to public libraries such as BiblioCommons or in add-in services like those from ChiliFresh (chilifresh.com/home/).

These features may not be effective or desirable in all circumstances, but the degree to which these features could be provided in ways that improve discoverability or that amplifies engagement with the library in other ways could be explored. More standardized APIs or interfaces between patron engagement or content enrichment platforms might help lower the threshold of difficulty in their implementation and result in more seamless integration.

d. Rich Media Materials and Collections

The current generation of library resource discovery products has been focused on textual materials and on text-oriented technologies. The initial phase of discovery centered on articles published in scholarly journals, theses, dissertations, books, and other texts. To the extent that audio and video materials are represented, they rely on the text of the transcripts for indexing, search, and retrieval. Textual representation of audio and video materials through transcriptions provides an important supplement to descriptive and administrative metadata to aid in the discovery of these materials.

Future discovery services may be able to offer search tools more able to exploit the visual content and qualities of video. Discovery services will benefit as libraries or content creators take advantage of automated video description tools to automatically index video through techniques such as speech-to-text or by mining closed caption tracks in addition to any manual processes for the creation of metadata. It will also be beneficial for discovery services to be empowered with specialized tools able to address the digital video or audio directly, through pattern matching, facial recognition, or other techniques that already exist or are emerging within specialized multi-media systems.

Libraries with discovery interfaces often continue to rely on separate platforms for the management and access of their digital collections. These collections often require specialized domain-specific search mechanisms and presentation methods relative to the media. Some types of content require page turning software, slideshow views, specialized image viewers, and the like. Discovery services have relatively weak support for the presentation and manipulation of images, video, digital recordings, and

other rich media objects. It is not a given that a generalized discovery service is the best way to provide access to these types of digital objects, but it warrants study of whether the access and interface capabilities can be improved to enable the possibility for those libraries that want to consolidate discovery of these materials. The impact of discovery services would certainly be enhanced as they become less exclusively oriented toward text and gain greater capabilities for the growing body of content based on other types of media.

e. Research Data Sets

Academic libraries have in recent years become more involved in the management of data sets produced through research projects. There are a variety of opportunities in expanding the involvement of discovery services into the realm of research data. It is important to facilitate the discovery of this data, especially for those interested in inspecting the data that corresponds to studies mentioned in the scholarly literature. There may also be some opportunity to include the research data itself at a more granular level within discovery, though this may involve many complications.

The involvement of libraries in this domain is nascent and the methodologies continue to evolve, so is not yet clear what the discovery needs are for these data sets. Some libraries may provide assistance to researchers as they develop grant application in the area of data management plans, some may be involved in the execution of those data management plans, and some may provide metadata repositories or preservation platforms to house that data. There seems to be significant potential benefit in standard metadata structures for describing research data sets and for making those available to institutional and global discovery environments. A key capability might include the ability to link to data sets from published articles based on that data, enabling other researchers to validate or replicate findings or to perform related studies based on that data.

A potential role for NISO might involve facilitating standards in the description of research data including linking mechanisms between data repositories and discovery interfaces.

f. Discovery and Access Related to Special Collections Materials

The current generation of discovery services does not necessarily provide adequate access into the specialized collections of the library, the archives of an institution (whether it be part of the library or a separate institution in the university), or in other departments that manage unique information resources.

Special collections and archives follow different concepts in the management of their collections and rely on a specialized set of metadata standards. Special collections tend to follow a more hierarchical approach to management and description. Depending on the collection's importance and the management resources available, there may be a single record for an entire collection, records created for each box, folder, or other grouping of materials, or for the individual items themselves, such as manuscripts, photographs, or physical objects within a collection. To provide better access to special collections, discovery services would need further development in supporting their metadata structures and hierarchical organizational concepts.

How discovery of specialized collections is distributed outside the local institution would also need to be addressed. Some collections would lend themselves to broad global discovery while others would necessarily need to be limited within the institution, to specific groups of users within an institution, or

to groups of individuals across institutions that might participate in an organization with interest or authority over that collection.

g. Analytics

Libraries and publishers have considerable interest in the ability to measure the performance of their discovery service and which resources are retrieved as a result of its use. The ways in which use of discovery services is recorded and evaluated needs to become more sophisticated. It should be possible, for example, to generate objective statistics that demonstrate the performance of the discovery environment relative to the content offerings of the library, revealing any skewing of results offered or of items selected by patrons toward or against any category of materials. These categories might include publishers, open access, scholarly or popular, or others. The use of a discovery service should be self-documenting relative to how it provides access to library collections.

Discovery services represent only a sub-set of overall use of the library's resources. The comprehensive use statistics can be gathered from publishers in COUNTER format and recorded and analyzed in electronic resource management systems.

The effectiveness of a discovery environment can also be measured by comparing its volume of use relative to the comprehensive statistics. This type of analysis can help assess the value of the discovery tool relative to the other paths in which library users gain access to library content materials.

h. Altmetrics

As alternative measures emerge relative to describing the impact of scholarly resources and the performance of academic libraries, to what extent can they become part of the discovery ecosystem? Can they be used in relevancy algorithms to help identify materials of higher interest or quality?

The work of the NISO Alternative Assessment Metrics (Altmetrics) Initiative will be an important resource to the advancement of altmetrics in many potential areas. Discovery services will likewise benefit from exploration of the topic and any recommendations it might make.

Some of the discovery services have already begun to incorporate altmetrics into their products or are in the process of developing related capabilities:

- Primo currently has an option to incorporate altmetric.com data.(initiatives.exlibrisgroup.com/2012/12/altmetrics-on-primo.html)
- EBSCO Information Services has acquired Plum Analytics, a company that specializes in altmetrics. (https://www.ebsco.com/news-center/press-releases/plum-analytics-becomes-part-of-ebsco-information-services)
- ProQuest has developed prototypes that incorporate altmetrics. (labs.proquest.com/projects/altmetrics)
- OCLC developed a partnership to incorporate WorldCat data into PlumX into prior to the acquisition of Plum Analytics by EBSCO. (https://www.oclc.org/news/releases/2013/201343dubllin.en.html)

6. Indexing Technologies: Harvesting Processes

The current slate of index-based discovery services rely on indexes populated with metadata and full text supplied primarily by commercial publishers through private agreements. The scope of the NISO Open Discovery Initiative included the technical mechanisms by which these data transfers take place. There was little interest expressed by either the content providers or the discovery service creators in prescribing a standard mechanism for data transfer. The current arrangements based on a variety of pragmatic exchange mechanisms or available standard protocols were deemed sufficient. As noted in the ODI Recommended Practice:

> The ODI recommends that the transfer of data from content providers to discovery service providers should make use of existing standards where applicable. Some of the standards and protocols most directly applicable include the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and KBART. ResourceSync also has strong potential as a mechanism for data transfer for discovery services once it is published by NISO and software tools become available for its implementation.

Transfer of content from publishers to discovery service creators remains in the domain of private commercial agreements. Proprietary content assets are considered strategic intellectual property that if disseminated freely beyond a publisher's own secure repositories or the controlled access through a discovery service could disrupt their business. As long as the current business models remain in place, the transfer of this content will remain constrained within these commercial agreements and any prescribed transfer mechanism will have little impact.

Open access content stands as one specific area that would greatly benefit from a standard exchange protocol for extraction of metadata and full-text content for the population of discovery services. Libraries that do not subscribe to a commercial discovery service may, for example, want to be able to build their own services based on open access materials available to their users both within and beyond their active subscriptions to content products. Commercial discovery services likewise may be interested in the ability to extract open access content from publishers regardless of whether there are formal agreements in place for proprietary content. As noted, NISO's *Access License and Indicators* recommended practice will support this need.

Any eventual transition to broad exposure of scholarly materials as linked data, including both proprietary and open access content, would naturally have a major impact in how discovery services are populated. Some services may want to cache or harvest the linked data into indexes or use other mechanisms to incorporate this knowledge into their discovery functionality.

7. Discovery beyond Library-Provided Interfaces

The recent study authored by Roger Schonfeld of Ithaka S+R [Schonfeld, Roger C. (2014).] highlighted the issue that much of the discovery process is not conducted through the interfaces provided by the library. While this report focuses primarily on issues related to library-oriented discovery services, any processes or technologies that can address other channels of discovery are also important to explore.

Discovery interfaces currently provide a very powerful tool for helping library patrons gain access to a very broad universe of content. Libraries rely on these tools to help direct users to resources that they have acquired on their behalf or that may be available to them through open access. The discovery services continue to improve in the content they address, in their ease of use, and in their sophistication of functionality. Nonetheless, library users in many cases do not go to the discovery tools provided by the library, but expect to find and gain access to library-provided content in other ways.

Google Scholar, which embodies many of the characteristics of a discovery service, stands as one of the most important channels through which students and scholars gain access to library-provided content. Libraries have very little influence on Google Scholar compared to the discovery services produced by the library-oriented vendors. Google was not a participant in the NISO Open Discovery Initiative, for example.

Inclusion of Google was not specifically addressed in the initial ODI Working Group. It would naturally be of interest to invite Google as a participant in any proposed new rounds of the Open Discovery Initiative. It is not clear if Google would provide a representative if invited, but given Google's position in the information discovery landscape, its participation would be valuable.

Other vectors of discovery and access to library resources take place though learning management systems. As previously discussed, integration with an LMS is a natural opportunity for discovery services. All of the discovery services already offer tools such as search widgets, reading list creation tools, and other mechanisms to assert the library's presence and its resources through the LMS.

Library patrons also often start their research with one of the general or specialized databases with which they are familiar. They may have these databases bookmarked in their browsers or connect through a link on the library's website. As patrons make use only of these narrower products, they will not be exposed to other materials that may be available from other sources. Having integrated mechanisms that transfer queries to the library's discovery service might be a feature that could help guide users to a broader range of resources. Many discovery services have the means to channel users into specialized resources, but there are few examples of tools that lead from specialized databases to general discovery results.

8. Open Discovery Initiative: Possible Next Steps

The initial phase of the NISO Open Discovery Initiative focused on establishing some expectations and issued a set of recommended practices relative to the transparency in how the indexes associated with discovery services are populated, as well as some very general expectations about the functionality of the retrieval of results. The ODI reflected an expectation that publishers of library-oriented content cooperate with discovery services by providing their content for indexing and were asked to disclose the extent to which they do so. Creators of discovery services were asked to disclose whether their products give any intentional bias or preference to content from specific organizations.

a. What Was Accomplished

The work of the ODI was an initial step to highlight some of the key issues in play in the realm of discovery services, to provide standard vocabulary for understanding the ecosystem, and to make some specific recommendations to improve transparency.

Key outcomes include the clear desire that libraries expect all of their content that they consider relevant to be represented with the discovery service of their choice. They expect discovery services to be neutral relative to the source of the content and for search results to be constructed without bias. Given uneven participation to date, one of the key recommendations is for content providers to disclose their participation with discovery service providers, with minimal and enhanced levels of compliance defined. Discovery service providers are prompted to provide assertions as to whether search results are delivered in a neutral way relative to content sources and business arrangements.

Coinciding with the final steps of ODI, there has been considerable positive movement in the participation of content providers with discovery services. Both ProQuest and EBSCO Information Services, which previously withheld participation with external discovery services, have either made partnerships or public announcements to make all or major portions of their materials available. (See: ProQuest is accelerating the interoperability of its products and services on all fronts; EBSCO Policy for Metadata Sharing & Collaboration with Discovery Service Vendors)

b. Focus Attention on A&I

The lack of participation by many A&I providers stands as one of the most critical problems in the current discovery services arena. Facilitating the resolution of this issue would provide great benefit to all the stakeholders defined in the discovery ecosystem.

The ODI recognized that content products based on proprietary abstracting and indexing services pose special challenges in the discovery ecosystem. A&I providers have strong interest in maintaining the value of their products and have concerns relative to how their content might be handled in a generalized discovery index. Many A&I providers continue to withhold their products from some or all of the discovery services.

This gap continues to be a concern for libraries. [Although ProQuest has recently begun working with Ex Libris and OCLC, the level of participation continues to be far from comprehensive. To date, no similar agreement has been announced between Ex Libris and EBSCO, for example. EBSCO has issued a set of public statements describing what content it will contribute to other discovery services, as noted above]

A&I products provide abstracts and metadata constructed by specialists in the field to provide the access points that specialists in the discipline require for comprehensive research within their fields. The metadata, for example, might include controlled vocabularies or ontologies that include related terms. This specialized metadata enables the retrieval of materials even when the query terms may not include the keywords that were used to index the article.

Due to the lack of participation of A&I creators, many of the discovery services aim to achieve the maximum coverage of the disciplines through other means. In the absence of A&I metadata, a discovery service may attempt to achieve similar functionality algorithmically using the full text of the journals covered by any given A&I provider. Discovery service creators that do not receive content from a specific A&I resource or aggregated database, tend to provide statements that describe that they include full text or metadata of a specified portion of that product, which they obtain from primary publishers or other sources. Libraries have to examine the coverage of a discovery service quite carefully to understand when a discipline-specific A&I database is included directly or whether it is covered indirectly through full-text or citation indexing.

More detailed disclosure for coverage of A&I resources

A&I resources remain an important component of the discovery ecosystem. Many library users continue to rely on the comprehensive and precise capabilities of these products within specialized disciplines, not necessarily feasible through a generalized discovery service.

One task of a follow-up ODI might be to address a more clear approach to describing how a discovery service covers A&I products. Statements such as: “We cover xx% of the titles in xx product” do not provide a candid disclosure that the specified titles are actually represented through content provided by other sources. Providing the actual metadata from the A&I products can improve the performance of discovery services, although the role of the A&I content may not necessarily be apparent to the users of the service.

The value of these A&I products persists despite the growing adoption of index-based discovery services. Concern remains, however, that sales of these products may erode if discovery services do not fully represent the capabilities of the A&I content or do not provide the means for end users or librarians to be aware when A&I content contributes to more relevant search results.

The basic tension between the interests of A&I providers and index-based discovery services remain largely unresolved through the initial phase of the Open Discovery Initiative. Another round of dialogue among these two stakeholder groups may be able to produce a recommended practice that addresses these concerns.

Activities that the new ODI workgroup might sponsor in this area could include an empirical study to assess the level of participation of A&I products across the current discovery services. Other tasks would include documenting the specific concerns of the A&I stakeholders regarding technical and business issues that currently stand as impediments to participation. Measures that would need to be taken in the technical operation of discovery services, business relationships, or other areas that would mitigate the factors that currently impede participation could be identified.

c. Data Exchange Protocols between Content Providers and Discovery Services

The initial ODI recommended practice did not result in any changes to the way that any of the technologies involved were developed or implemented. Rather it focused on data exchange issues. Even in the area of how content providers provide data and metadata to discovery service providers, there was little interest expressed in new protocols or techniques to support the interchange. The senders and recipients generally acknowledge that they are able to use private exchange mechanisms or existing data formats or protocols to efficiently make these exchanges.

Any future rounds of discussions may need to delve deeper in to the mechanics of discovery services. We can anticipate that the concerns expressed by A&I providers may not be able to be addressed through disclosure agreements, but rather may require structural changes in the way that their data are indexed and presented. These changes may or may not be consistent with the internal workings of some discovery services. While public pressure informed by the disclosure requirements of the ODI recommended practice may prod some A&I providers into cooperation with discovery service providers, more technical or structural changes may be needed to enable fuller participation among these two stakeholder groups.

d. Relevancy

The initial ODI recommended practices also did not address how discovery services perform relative to the identification of result candidates and in the calculation of relevancy order. Rather, discovery service providers are asked to disclose whether or not they determine relevancy in ways that might prejudice search results relative to any given publisher or provider. Technologies and techniques each discovery service employs are considered proprietary and are an important point of competitive differentiation.

Several of the comments received to the drafts of the recommended practice expressed concern that this topic was considered out of scope. A next-round task of the ODI could be to take a closer look at the area of relevancy and work to find a more substantive vehicle for assuring libraries of the objectivity of relevancy beyond a general affirmative statement. The challenge lies in probing into the functionality of these discovery services to make their performance more transparent without intruding into the proprietary internal mechanisms that discovery service providers may consider as competitive advantages.

e. Interoperability with Resource Management Systems

The initial ODI recommended practice did not focus on the relationship between discovery service providers and the providers of resource management systems. An important part of a library's technical infrastructure lies in the intersection between these two categories of components. The provider of a resource management system, such as an integrated library system, electronic resource management systems, library services platform, or digital asset management system, may or may not include a discovery interface or a central discovery index.

The next round of ODI could explore whether or not libraries have the same expectations for interoperability among discovery services as they do in the realm of publisher participation in discovery indexes.

In the same way that some natural tension and concern applies when a discovery service creator is also a content provider, similar concerns apply when the discovery service creator also produces a resource management system. There is a natural affinity for bundling resource management systems and discovery services. Benefits may include leveraging common knowledge bases, similarities in conceptual design, and ability to tightly integrate the products through proprietary mechanisms or a highly optimized suite of APIs.

The next phase of ODI might address the question: When an organization produces both a resource management system and a discovery service, to what extent should that resource management system also be interoperable with other discovery services?

Developers that produce resource management systems but do not produce their own index-based discovery service may have an interest in the ability to interoperate with any of the index-based discovery services. This interoperability may include full replacement of the discovery layer, including both the patron interface and the functionality provided through the central index, or it may involve using the interface provided by the developer of the resource management system with the functionality of the index-based discovery service. Another scenario involves libraries that may prefer to use their own discovery interface, such as one based on VuFind or Blacklight, which will work with their choice of resource management systems and index-based discovery services.

This area of interoperability between resource management systems and discovery services has been addressed to a certain extent in the ILS-DI [See Breeding, Marshall. (2008).] These mechanisms were conceived prior to the advent of index-based discovery services. The areas addressed include how metadata from an ILS are extracted and synchronized to populate a discovery index, how the current status and availability of items is expressed, and how tasks oriented to end users are achieved such as placing requests, viewing items charged, and viewing or changing details from a patron record resident in an ILS through the discovery interface. The ILS-DI does not address whether or how an index-based discovery service makes its functionality available to other discovery interfaces or resource management systems.

The ILS-DI was an initiative of the Digital Library Federation before it became a program of CLIR. While the ILS-DI work stands as a reference model, it has not been maintained as an active specification with ongoing maintenance. The interactions and interoperability between discovery interfaces, resource management systems, and other related systems has become increasing complex since that effort ended. The new phase of ODI might make recommendations regarding whether a new initiative similar to ILS-DI focused specifically on these interoperability issues is warranted. Work in this area might address whether a discovery service provider cooperates with each of the applicable resource management products and to explore whether any common set of techniques might be defined to facilitate interoperability. Should there, for example, be a common set of APIs that define the interoperability among resource management systems, discovery interfaces, and index-based discovery services?

f. Summary of Possible Activities for ODI Phase 2

A potential second phase of the Open Discovery Initiative might extend the work of the initial recommended practice to address the following topics:

- Concerns of A&I providers with the goal of wider participation between A&I providers and discovery service creators
- More detailed disclosures regarding relevancy calculations in discovery services to strengthen confidence that results are not preferential for or against any category of content
- Interoperability between discovery service providers and other related components, including discovery interfaces, resource management systems, learning management systems, or other relevant external systems.

9. Potential Areas of Action for NISO

Based on the information gathered and trends underway, there are areas in which NISO could facilitate progress in the advancement of discovery services and related technologies. Gaps remain in the expected capabilities of these products and there are important areas of development yet to be exploited. The genre of discovery services, dominated by four major competing products, continues to see ambitious development and rapid adoption by libraries. This niche of the library technology industry cannot be considered unhealthy. But given its strategic importance to libraries, it warrants ongoing study to identify areas where the various stakeholder groups can benefit through new cooperative activities or through refinement of practices.

In this vein, there are some areas where NISO can play a beneficial role through extensions of some of its existing workgroups or programs. Some of these actions might include:

- **Convene a second phase of the Open Discovery Initiative** (see 8.f above). This second round of the initiative would focus on working through the areas of tension between the A&I providers and discovery service providers. Resolution of some of these issues could potentially result in greater participation by the A&I providers, one of the most pressing issues in the discovery arena. A second priority would delve deeper into how discovery services identify result candidates and perform relevancy ordering to give libraries more confidence that these services operate in a neutral way relative to the source of the content. The issue of the interoperability between discovery services and resource management systems should also be addressed.
- Launch a study group or research project focused on **open linked data** and opportunities to facilitate the exposure of metadata in the scholarly publishing arena to increase new avenues of discovery that extend the capabilities of index-based discovery services.
- NISO currently has a workgroup charged with analyzing the impact of alternative metrics in measuring the use and impact of information resources. One additional thread of that work could address how **Altmetrics** can be incorporated into the discovery services ecosystem to improve relevancy or other areas of their performance.
- In support of discoverability of resources without library-provided discovery tools, a workgroup could explore recommended practices related to the **presentation of content on the web in ways that maximize exposure and indexing by Google Scholar** or other search tools. There seems to be momentum toward the use of schema.org and other techniques for encoding documents to enhance discoverability. OCLC, Zepheira, and others are already active in this area, but there may be room for NISO to extend these efforts.

10. Longer Term Prospects for Discovery

Much of this white paper has been focused on opportunities to make incremental improvements in the area of library resource discovery in the relatively short term. Extending a view to the longer horizon, there may be more systemic changes that would drive greater substantive change in this arena. The current generation of discovery services has been shaped by the pervading realities in scholarly publishing, technology, information architectures, and a specific set of vendors in the industry.

More radical change becomes possible, and necessary, when some or all of these broad environmental factors change. A dramatic shift toward open access would enable opportunities for discovery of this material more dramatically than the gradual evolution seen so far. A universe of scholarly content dominated by open access could potentially lower the threshold of difficulty in the creation of discovery services.

Likewise, a comprehensive exposure of scholarly material as linked data would have the potential to undermine the prevailing model of index-based discovery to one of dynamic linking or exploration.

It is also possible that business events might prompt more sudden turns in the direction of resource discovery. Abrupt consolidation or elimination of one or more of the current products that narrow the field of options to an intolerable level could prompt new initiatives for open access and open source alternatives that until now have not proven themselves necessary. The withdrawal of Google Scholar or the introduction of a new service of even greater capability would also disrupt the landscape.

Each of these larger-scale changes would at a minimum accelerate some of the possible areas of change or improvement in the discovery services arena discussed in this paper. Dramatic change seems unlikely, though. Each of the stakeholder groups involved in discovery—content providers, discovery service creators, and libraries—have long been adopters of incremental change more than abrupt turning points. This view of history leads towards expectations of pragmatic efforts that build on the current set of realities rather than a visionary approach based on new possibilities in the publishing arena or on technological breakthroughs.

11. Literature and Bibliography

The following resources provide relevant information regarding the resource discovery arena, including articles, book chapters, and white papers that address relevant topics, and recent studies on the performance of discovery service implementations or those that address conceptual issues. Selected resources were reviewed in section B.

Adams, Jonathan. (2012). “Collaborations: the rise of research networks.” *Nature*, 490: 335–336. Available at: http://dx.doi.org/10.1038/490335a

Arlitsch, Kenning, Patrick Obrien, Jason A. Clark, Scott W. H. Young, and Doralyn Rossmann. (2014). “Demonstrating Library Value at Network Scale: Leveraging the Semantic Web with New Knowledge Work.” *Journal of Library Administration,* 54 (5): 413-425. Available at: http://dx.doi.org/10.1080/01930826.2014.946778

Asher, Andrew D., Lynda M. Duke, and Suzanne Wilson. (2013). “Paths of Discovery: Comparing the Search Effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and Conventional Library Resources.” *College and Research Libraries*, 74 (5): 464-488. Available at: http://dx.doi.org/ 10.5860/crl-374

Balnaves, Edmund. (2013). “From OPAC to Archive: integrated discovery and digital libraries with open source.” Presented at the *79th IFLA General Conference and Assembly, August 17-23, 2013, Singapore*. Available at: http://library.ifla.org/79/1/108-balnaves-en.pdf

Belford, Rebecca. (2014). “Evaluating Library Discovery Tools through a Music Lens.” *Library Resources & Technical Services*, 51(1), 49–72. Available at: http://dx.doi.org/10.5860/lrts.58n1.49

Breeding, Marshall. (2008). “Progress on the DLF ILS Discovery Interface API: The Berkeley Accord.” *Information Standards Quarterly,* 20 (3): 18-19. Available at: http://www.niso.org/apps/group_public/download.php/5637/ISQv20no3.pdf

Breeding, Marshall. (2012). Knowledge Base and Link Resolver Study: General Findings. LIBRIS nationella biblioteksystem. Available online: http://www.kb.se/dokument/Knowledgebase_linkresolver_study.pdf

Breeding, Marshall. (2013). "Next-Generation Discovery: An Overview of the European Scene." In: Chambers, Sally, ed. *Catalog 2.0: The Future of the Library Catalog.* Chicago: ALA Neal-Schuman, pp. 37-64. ISBN: 978-1-55570-943-3

Breeding, Marshall. (2014). “Library Resource Discovery Products: Context, Library Perspectives, and Vendor Positions.” *Library Technology Reports,* 50 (1). Available at: http://dx.doi.org/10.5860/ltr.50n1

Calvert, Kristin. (2015). “Maximizing Academic Library Collections: Measuring Changes In Use Patterns Owing to EBSCO Discovery Service.” *College and Research Libraries*, 76 (1): 81-99. Available online: http://dx.doi.org/10.5860/crl.76.1.81

Chickering, F. William, and Sharon Q. Yang. (2014). “Evaluation and Comparison of Discovery Tools: An Update.” *Information Technology and Libraries,* 33 (2): 5-30. Available at: http://dx.doi.org/10.6017/ital.v33i2.3471

Elleroa, Nadine P. (2014). "Exploring Library Discovery Positions: Are They Emerging or Converging?" *Journal of Web Librarianship*, 8 (4): 331-348. Available at: http://dx.doi.org/10.1080/19322909.2014.963778

Hoeppner, Athena. (2012). "The Ins and Outs of Evaluating Web-Scale Discovery Services." *Computers in Libraries*, 32 (3): 6-11. Available at: http://www.infotoday.com/cilmag/apr12/Hoeppner-Web-Scale-Discovery-Services.shtml

Kabashi, Art; Christine Peterson; Tim Prather. (2014). *Discovery Services: A White Paper for the Texas State Library and Archives Commission*. Austin, TX: Texas State Library & Archives Commission. Available at: https://www.tsl.texas.gov/sites/default/files/public/tslac/lot/TSLAC_WP_discovery__final_TSLAC_20140912.pdf

Kelley, Michael. (2012). "Stakeholders Strive to Define Standards for Web-Scale Discovery Systems." *The Digital Shif*t (*Library Journal / School Library Journal)*. Available at: http://www.thedigitalshift.com/2012/10/discovery/coming-into-focus-web-scale-discovery-services-face-growing-need-for-best-practices/

Kemperman, Suzanne Saskia. (2014). *Success Strategies for Electronic Content Discovery and Access: A Cross-Industry White Paper*. Dublin, OH: OCLC, Inc. Available at: http://www.oclc.org/content/dam/oclc/reports/data-quality/215233-SuccessStrategies.pdf

Kortekaas, Simone. "Thinking the Unthinkable: a Library without a Catalogue – Reconsidering the Future of Discovery Tools for Utrecht University library." Presented at the *LIBER General Annual Conference 2012*. Available at: http://www.libereurope.eu/blog/thinking-the-unthinkable-a-library-without-a-catalogue-reconsidering-the-future-of-discovery-to
Note: This post is a summary of the author's presentation at the LIBER General Annual Conference 2012. A video of the presentation is available at http://uttv.ee/naita?id=12538

Levine-Clark, Michael. (2014). "Access to Everything: Building the Future Academic Library Collection." *portal: Libraries and the Academy,* 14 (3): 425-437. Available at: http://dx.doi.org/10.1353/pla.2014.0015

Levine-Clark, Michael, Jason Price, and John McDonald. (2014). "Discovery or Displacement? A Large-Scale Longitudinal Study of the Effect of Discovery Systems on Online Journal Usage." In: Beth R. Bernhardt, Leah H. Hinds, and Katina P. Stauch, eds. *Too Much is Not Enough, Charleston Conference Proceedings, 2013*. Purdue University Press, 2014. Available at: http://dx.doi.org/10.5703/1288284315331

Lundrigan, Courtney, Kevin Manuel, and May Yan. (2015). "Pretty Rad: Explorations in User Satisfaction with a Discovery Layer at Ryerson University." *College and Research Libraries,* 76 (1): 43-62. Available at: http://dx.doi.org/ 10.5860/crl.76.1.43

NFAIS. (2013). *Recommended Practices: Discovery Services*. Philadelphia, PA: National Federation of Advanced Information Services. Available at: https://nfais.memberclicks.net/assets/docs/BestPractices/recommended_practices_final_aug_2013.pdf

NISO. (2014). *Open Discovery Initiative: Promoting Transparency in Discovery*. NISO RP-19-2014. Baltimore, MD: National Information Standards Organization. Available at: http://www.niso.org/publications/rp/rp-19-2014

Orduña-Malea, Enrique, Juan Manuel Ayllón, Alberto Martín-Martín, and Emilio Delgado López-Cózar. (2014). About the Size of Google Scholar: Playing the Numbers. Granada: EC3 Working Papers No. 18. Available at: http://arxiv.org/abs/1407.6239

Sadeh, Tamar. (2013). "From Search to Discovery." Presented at the *79th IFLA General Conference and Assembly, August 17-23, 2013, Singapore*. Available at: http://library.ifla.org/104/1/098-sadeh-en.pdf

Schilling, Virginia. (2013). Transforming Library Metadata into Linked Library Data: Introduction and Review of Linked Data for the Library Community, 2003–2011. Document created for ALCTS CaMMS Research and Publications Committee. Chicago: American Library Association. Available at: http://www.ala.org/alcts/resources/org/cat/research/linked-data

Schonfeld, Roger C. (2014). *Does Discovery Still Happen in the Library? Roles and Strategies for a Shifting Reality*. New York: Ithaka S+R. Available at: http://www.sr.ithaka.org/sites/default/files/files/SR_Briefing_Discovery_20140924_0.pdf

Somerville, Mary M., and Lettie Y. Conrad. (2014). C*ollaborative Improvements in the Discoverability of Scholarly Content: Accomplishments, Aspirations, and Opportunities*. A SAGE White Paper. Thousand Oaks, CA: SAGE Publications. Available at: http://www.sagepub.com/repository/binaries/pdf/improvementsindiscoverability.pdf

Tay, Aaron. (2013). "How are Discovery Systems Similar to Google? How Are They Different?" *Musings about Librarianship* [blog]. Available at: http://musingsaboutlibrarianship.blogspot.sg/2013/04/how-are-discovery-systems-similar-to.html

Tay, Aaron. (2014). "8 Surprising Things I Learnt about Google Scholar." *Musings about Librarianship* [blog]. Available at: http://musingsaboutlibrarianship.blogspot.com/2014/06/8-surprising-things-i-learnt-about.html

Van Noorden, Richard. (2014). "Google Scholar Pioneer on Search Engine's Future." *Nature*. Available at: http://dx.doi.org/10.1038/nature.2014.16269

Yunta, Luis Rodríguez. (2015). "Servicios de descubrimiento basados en un índice centralizado: su expansión en las bibliotecas académicas españolas y futuras líneas de investigación [Discovery services based on a centralized index: its expansion in the Spanish and future lines of research academic libraries]." Posting to *IWETEL listserv*. https://listserv.rediris.es/cgi-bin/wa?A2=ind1501B&L=IWETEL&F&S&P=15938