

INFORMATION STANDARDS QUARTERLY
SPRING 2015 | VOL 27 | ISSUE 1 | ISSN 1041-0031

ISQ

TOPIC

YEAR IN REVIEW AND STATE OF THE STANDARDS

NISO 2014 YEAR IN REVIEW

TC46 2014 YEAR IN REVIEW

THE FUTURE OF LIBRARY RESOURCE DISCOVERY

STATE OF THE STANDARDS



Marshall
Breeding

MARSHALL BREEDING

The Future of Library Resource Discovery

NISO's Discovery to Delivery Topic Committee commissioned a white paper on *The Future of Library Resource Discovery* from Marshall Breeding as part of its ongoing examination of areas in the discovery landscape that the information community could potentially standardize. The white paper was published in February 2015. This article provides an extracted summary of the paper. The full paper is available for download from the NISO website.

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 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.

Categories of Discovery Systems

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

includes the discovery interface, end-user interface, interoperability with a link resolver, local search and retrieval, ability to interactively communicate with the library's ILS implementation, and access to remote index platforms via API.

- » Commercial examples: Ex Libris® Primo®, SirsiDynix® Enterprise®, BiblioCommons BiblioCore, ProQuest® AquaBrowser® Library, Innovative Interfaces Encore
- » Open Source examples: Blacklight, VuFind, eXtensible Catalog, Franklin

➤ 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.

- » Commercial examples: Primo® and Primo Central from Ex Libris Group, EBSCO Discovery Service™ from EBSCO Information Services, Summon® from ProQuest, WorldCat® Discovery Service from OCLC®

➤ **Local index content**

is the ability to incorporate local resources in addition to article-level scholarly content from proprietary and open access sources. Some of these categories of content include archival material, digital collections, institutional repositories and electronic theses and dissertations, and museum or exhibition materials.

➤ **Non-library discovery 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.

➤ **Article-level discovery services**

not based on central indexes such as federated search are still in use. While there has been a major shift toward reliance on central indexes in support of discovery, the change is not universal.

➤ **Public library discovery services**

need 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. 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.

» Commercial examples: BiblioCommons, AquaBrowser, ProPAC for Polaris, Encore from Innovative Interfaces and the LS2 PAC from The Library Corporation

➤ **Comprehensive library portals**

include a discovery component that resides among other parts of a library's overall web presence.

» Commercial examples: Iguana from Infor Library Solutions, Arena from Axiell, BiblioCMS from BiblioCommons, and Enterprise from SirsiDynix (optional capability)

Standards and Recommended Practices

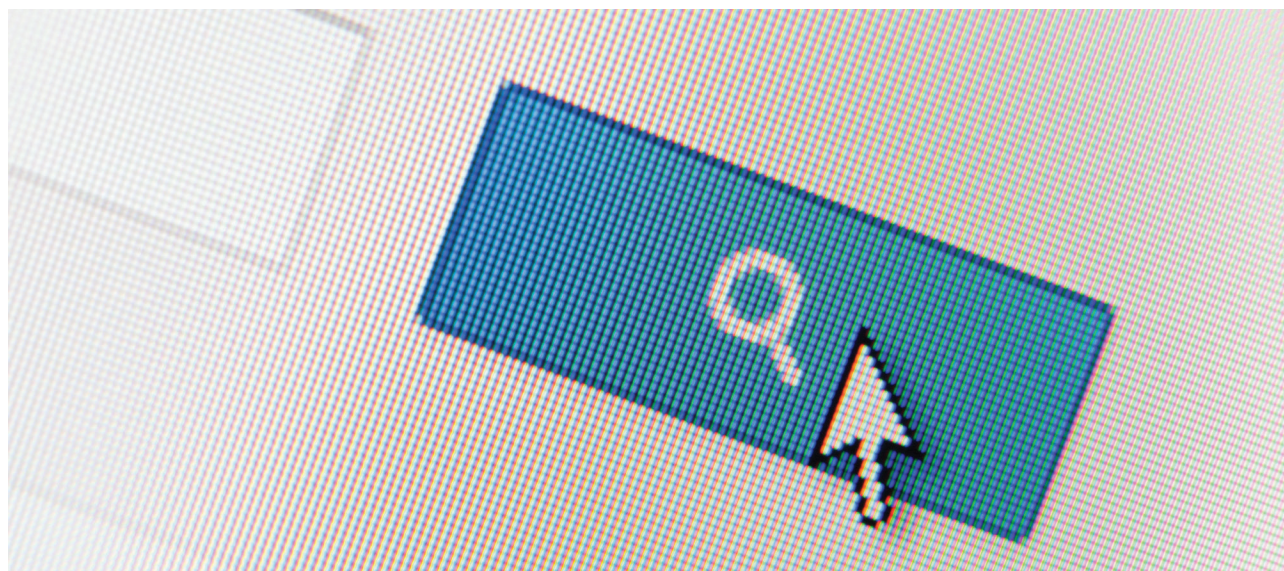
Discovery-specific project initiatives include:

➤ **NISO Open Discovery Initiative**

The NISO Open Discovery Initiative Working Group developed a recommended practices document for *Promoting Transparency in Discovery* (NISO RP-19-2014).

➤ **NFAIS Recommended Practices: Discovery Services**

The National Federation of Advanced Information Services' *Recommended Practices: Discovery Services* was published



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.”

CONTINUED »

in August 2013 to address the interests of the providers of abstracting and indexing services.

➤ **Discovery: A metadata ecology for UK Education and Research**

This Discovery initiative in the United Kingdom was active between 2011 and 2012 with the intent to improve discovery of resources through improved metadata practices. One of the outcomes of the project was the development of a set of *Discovery Open Metadata Principles*.

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.

➤ **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 currently accomplished through entirely proprietary methods. A variety of application programming interfaces (APIs) are involved in the discovery services ecosystem, but there has been little progress toward developing commonality. How an index-based discovery service interacts with discovery interfaces also lacks standardization.

Open Access Global Discovery Service or Index

The index-based discovery service arena remains entirely dominated by commercial providers. 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. A few open source discovery interfaces are currently available to provide the end-user interface. 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.



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.

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. 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. 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. 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. Any project to create an open

access discovery interface would need an extensive network of participants to initiate discussions and coordinate content contributions from publishers.

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. 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. To enable full interoperability, additional APIs or other mechanisms must be enabled in the ILS.

Many vendors of resource management systems (including both integrated library systems and library services platforms) also offer discovery services. 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. Many libraries may also prefer to assemble their own discovery environment based on an open source tool.

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.

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.

Gap Analysis

What are some of the features or concepts that may not be fully realized in the current generation of discovery services?

➔ 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. 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. The general participation of A&I resources in the discovery services arena remains moderate to weak.

➔ Internationalization and Multi-Lingual Coverage

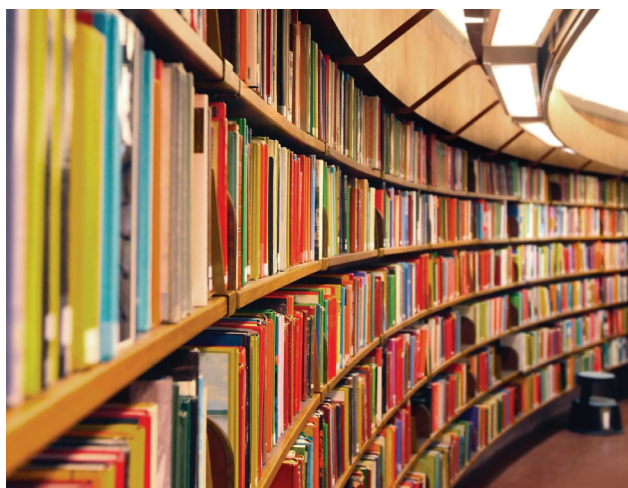
Coverage of bibliographic resources from diverse international sources is growing. Coverage of article-level scholarly resources, primary research resources, and other material in non-English languages is likewise improving, but is far from universal. The content represented in discovery indexes is becoming increasingly heterogeneous by language, which introduces challenges in search and retrieval. Cross-language searching remains fertile ground for future development in discovery services.

➔ Coverage of Open Access Materials

Each of the discovery services includes commercially published open access titles, materials from major disciplinary open access servers, and can tap into open access materials through centralized services such as OAlster. The challenge is how to expose open access materials from a variety of sources. 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 services.

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.

CONTINUED »



Online catalogs excelled at providing precise methods for interacting with a library's local collections, enabling patrons to browse through collections based on name or subject authority databases, and to virtually browse items as they would be shelved based on call number indexes.

➔ Precision and Known-Item Searching

Advanced and precision searching continue to be areas of interest for discovery services. Online catalogs excelled at providing precise methods for interacting with a library's local collections, enabling 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. The authority work performed 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. 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 Rankings

The way in which discovery services order search results is critical. Relevance ranking remains one of the key issues that impedes support of librarians for these products,

and improving relevancy ranking has been a high priority for the developers of discovery services. However, how relevancy functions within each of the discovery services remains in the proprietary realm and is considered one of the main competitive features. Expectations for transparency in how discovery services calculate relevancy could be a positive factor in improving the performance and the acceptance of these products.

➔ Enhanced Discoverability through Non-Textual Associations

Discovery services aim to provide enhanced discovery beyond keyword matching. They may 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. Though progress has been made, discovery services still have much room for improvement.

➔ 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. 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.

➔ Learning Management Systems

It is critical for the content and functionality of discovery 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. The ability for instructors to identify reading materials held by the library for a course represents a significant opportunity. 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.

Opportunities for Future Enhancements in Discovery Services

The genre of discovery services will continue to be enhanced to add new functionality and capabilities in response to requests from libraries and to improve their commercially competitive position. Some useful improvements and enhancements might be:

➤ Application Programming Interfaces

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.

➤ Expanding API Ecosystem

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.

➤ Social Features

Many libraries are interested in enabling individuals to interact with their collections in a variety of ways. For example, 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. 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.

➤ Rich Media Materials and Collections

The current generation of library resource discovery products has been focused on textual materials and on text-oriented technologies. To the extent that audio and video materials are represented, they rely on the text of the transcripts for indexing, search, and retrieval. Future discovery services may be able to offer search tools more able to exploit the visual content and qualities of video by taking advantage of automated video description tools to automatically index video. 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.



Many libraries are interested in enabling individuals to interact with their collections in a variety of ways. 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.

CONTINUED »

➔ Research Data Sets

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 and there may also be some opportunity to include the research data itself at a more granular level within discovery. 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.

➔ 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, or unique information resources in other departments. Special collections and archives follow different concepts in the management of their collections, rely on a specialized set of metadata standards, and tend to follow a more hierarchical approach to management and description. To provide better access to special collections, discovery services would need further development in supporting their metadata structures and hierarchical organizational concepts.

➔ 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.

➔ 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?

Potential Areas of Action for NISO

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 to address the tasks proposed in the white paper.

- » Launch a study group or research project focused on open linked data and opportunities to facilitate the exposure of metadata in index-based discovery services.
- » NISO's Alternative Assessment Metrics Initiative could investigate how altmetrics can be incorporated into the discovery services ecosystem to improve relevancy or other areas of their performance.
- » 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.

I NR | doi: 10.3789/isqv27n01.2015.04

MARSHALL BREEDING (marshall.breeding@librarytechnology.org) is an independent consultant, speaker, and author, specializing in the areas of library management and end-user discovery and service delivery. He is the founder and editor of Library Technology Guides (<http://librarytechnology.org/>) and was Co-chair of NISO's Open Discovery Initiative.

The Future of Library Resource Discovery white paper

http://www.niso.org/publications/white_papers/discovery/

NFAIS Recommended Practices: Discovery Services

https://nfais.memberclicks.net/assets/docs/BestPractices/recommended_practices_final_aug_2013.pdf

NISO Access License and Indicators Recommended Practice

<http://www.niso.org/publications/rp/rp-22-2015/>

NISO KBART (Knowledge Bases and Related Tools) Recommended Practice

<http://www.niso.org/publications/rp/rp-9-2014/>

NISO Open Discovery Initiative

<http://www.niso.org/workrooms/odi/>

NISO ResourceSync standard

<http://www.niso.org/standards/z39-99-2014/>

OAI-Protocol for Metadata Harvesting (PMH)

<http://www.openarchives.org/pmh/>

UK Discovery initiative

<http://discovery.ac.uk/>



RELEVANT
LINKS